1-1-2016

Statistical Methods for Time-Conditional Survival Probability and Equally Spaced Count Data

Victoria Alexandria Gamerman
University of Pennsylvania, victoria.gamerman@gmail.com

Statistical Methods for Time-Conditional Survival Probability and Equally Spaced Count Data

Abstract
This dissertation develops statistical methods for time-conditional survival probability and for equally spaced count data. Time-conditional survival probabilities are an alternative measure of future survival by accounting for time elapsed from diagnosis and are estimated as a ratio of survival probabilities. In Chapter 2, we derive the asymptotic distribution of a vector of nonparametric estimators and use weighted least squares methodology for the analysis of time-conditional survival probabilities. We show that the proposed test statistics for evaluating the relationship between time-conditional survival probabilities and additional time survived have central Chi-Square distributions under the null hypotheses. Further, we conducted simulation studies to assess the empirical probability of making a type I error for one of the hypotheses tests developed and to assess the power of the various models and statistics proposed. Additionally, we used weighted least squares techniques to fit regression models for the log time-conditional survival probabilities as a function of time survived after diagnosis to address clinically relevant questions. In Chapter 3, we derive the asymptotic distribution of time-conditional survival probability estimators from a Weibull parametric regression model and from a Logistic-Weibull cure model, adjusting for continuous covariates. We implement the weighted least squares methodology to assess relevant hypotheses. We create a statistical framework for investigating time-conditional survival probability by developing additional methodological approaches to address the relationship between estimated time-conditional survival probabilities, time survived, and patient prognostic factors. Over-dispersed count data are often encountered in longitudinal studies. In Chapter 4, we implement a maximum-likelihood based method for the analysis of equally spaced longitudinal count data with over-dispersion. The key features of this approach are first-order antedependence and linearity of the conditional expectations. We also assume a Markovian model of first order, implying that the value of an outcome on a subject at a specific measurement occasion only depends on the value at the previous measurement occasion. Our maximum likelihood approach using the Poisson model for count data benefits from a simple interpretation of regression parameters, like that in GEE analysis of count data.

Degree Type
Dissertation

Degree Name
Doctor of Philosophy (PhD)

Graduate Group
Epidemiology & Biostatistics

First Advisor
Phyllis A. Gimotty

Second Advisor
Justine Shults

STATISTICAL METHODS FOR TIME-CONDITIONAL SURVIVAL PROBABILITY AND EQUALLY SPACED COUNT DATA

Victoria A. Gamerman

A DISSERTATION

in

Epidemiology and Biostatistics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2016

Supervisor of Dissertation

Phyllis A. Gimotty

Professor of Biostatistics

Co-Supervisor of Dissertation

Justine Shults

Professor of Biostatistics

Graduate Group Chairperson

John H. Holmes, Professor of Medical Informatics in Epidemiology

Dissertation Committee

Susan Ellenberg, Professor of Biostatistics

DuPont Guerry, IV, Emeritus Professor of Medicine

STATISTICAL METHODS FOR TIME-CONDITIONAL SURVIVAL PROBABILITY AND EQUALLY SPACED COUNT DATA

ACKNOWLEDGEMENT

I would like to acknowledge the support of my committee and thank the Biostatistics faculty and my fellow students.

ABSTRACT

STATISTICAL METHODS FOR TIME-CONDITIONAL SURVIVAL PROBABILITY AND EQUALLY SPACED COUNT DATA

Victoria A. Gamerman

Phyllis A. Gimotty

Justine Shults

This dissertation develops statistical methods for time-conditional survival probability and for equally spaced count data. Time-conditional survival probabilities are an alternative measure of future survival by accounting for time elapsed from diagnosis and are estimated as a ratio of survival probabilities. In Chapter 2, we derive the asymptotic distribution of a vector of nonparametric estimators and use weighted least squares methodology for the analysis of time-conditional survival probabilities. We show that the proposed test statistics for evaluating the relationship between time-conditional survival probabilities and additional time survived have central χ^2-distributions under the null hypotheses. Further, we conducted simulation studies to assess the empirical probability of making a type I error for one of the hypotheses tests developed and to assess the power of the various models and statistics proposed. Additionally, we used weighted least squares techniques to fit regression models for the log time-conditional survival probabilities as a function of time survived after diagnosis to address clinically relevant questions. In Chapter 3, we derive the asymptotic distribution of time-conditional survival probability estimators from a Weibull parametric regression model and from a Logistic-Weibull cure model, adjusting for continuous covariates. We implement the weighted least squares methodology to assess relevant hypotheses. We create a statistical framework for investigating time-conditional survival probability by developing additional methodological approaches to address the relationship between estimated time-conditional survival probabilities, time survived, and patient prognostic factors. Over-dispersed count data are often encountered in longitudinal studies. In Chapter 4, we implement a maximum-likelihood based method for the analysis of equally spaced longitudinal count data with over-dispersion. The key features of this approach are first-order antedependence and linearity of the conditional expectations. We also assume a Markovian model of first order, implying that the value of an outcome on a subject at a

specific measurement occasion only depends on the value at the previous measurement occasion. Our maximum likelihood approach using the Poisson model for count data benefits from a simple interpretation of regression parameters, like that in GEE analysis of count data.

TABLE OF CONTENTS

ACKNOWLEDGEMENT iii

ABSTRACT iv

LIST OF TABLES viii

LIST OF ILLUSTRATIONS ix

CHAPTER 1 : INTRODUCTION 1

1.1 Time-Conditional Survival Probability Methods 1

1.2 Longitudinal Count Data Methods 7

1.3 Dissertation Structure 9

CHAPTER 2 : NONPARAMETRIC TIME-CONDITIONAL SURVIVAL PROBABILITY 10

2.1 Introduction 10

2.2 Estimation of Time-Conditional Survival Probabilities 12

2.3 Distribution Theory 13

2.4 Hypothesis Testing Using Weighted Least Squares 17

2.5 Simulation Studies 28

2.6 Application: Staging Procedure and Time-Conditional Survival Probability for Stage II Melanoma Patients 34

2.7 Discussion 48

CHAPTER 3 : PARAMETRIC TIME-CONDITIONAL SURVIVAL PROBABILITY 55

3.1 Introduction 55

3.2 Parametric Time-Conditional Survival 55

3.3 An Example: The Weibull Distribution 65

3.4 Application to Real-World Data 69

3.5 Discussion 84

CHAPTER 4 : Analysis of Longitudinal Count Data with Specified Marginal Means and First-order Antedependence 93
4.1 Introduction 93
4.2 Methods 95
4.3 Application 99
4.4 Simulation Studies 103
4.5 Discussion 105
CHAPTER 5 : Discussion 114
APPENDICES 118
BIBLIOGRAPHY 173

LIST OF TABLES

TABLE 2.1 : SEER melanoma contrasts of profile-based differences 51
TABLE 2.2 : Estimates of log time-conditional melanoma-specific survival probabilities, their variance-covariance matrix, and estimates from the saturated (H_1), QM, LM, and GM models 52
TABLE 2.3 : Parameter estimates for the multivariable analysis 53
TABLE 2.4 : Estimates of time-conditional melanoma-specific survival probabilities from the multivariable analysis 54

TABLE 3.1 : Maximum likelihood estimates of the Weibull survival distribution for disease-specific survival of esophageal cancer patients adjusting for tumor length. . 88
TABLE 3.2 : Maximum likelihood estimates of the Weibull mixture cure model for disease-specific survival. 89
TABLE 3.3 : Estimated covariance matrix* for the time-conditional survival probability from the Weibull mixture cure model for disease-specific survival. 90
TABLE 3.4 : Estimates of 5-year time-conditional survival probability from the Weibull mixture cure model for disease-specific survival adjusting for fixed gender (male), fixed age at diagnosis (60 years), fixed tumor thickness (3.58 mm) and varying staging type (clinical versus pathological), number of nodes examined, and ulceration status along with the estimated covariance and correlation matrices for the alternative hypothesis. 91
TABLE 3.5 : Change in 5-year time-conditional survival probability given 1 and given 10 years after diagnosis from the Weibull mixture cure model for disease-specific survival with Bonferroni adjustment. 92

TABLE 4.1 : Estimated parameters from the ML, GEE, and Poisson models in the analysis of the doctor visits data. 107
TABLE 4.2 : Mean and variance for the placebo and treatment groups. 108
TABLE 4.3 : Estimated parameters from the GEE and ML approaches for analysis of the epilepsy data when Period is included in the models. 109
TABLE 4.4 : Estimated parameters from the GEE and ML approaches for analysis of the epilepsy data when Period is not included in the models. 110
TABLE 4.5 : Small sample efficiencies for evaluating the AR(1) correlation structure for varying values of α and sample size per group. 111
TABLE 4.6 : Percent bias for evaluating the AR(1) correlation structure for varying values of α and sample size per group. 112
TABLE 4.7 : Coverage probabilities for the ML and GEE approaches with the AR(1) correlation structure for varying values of α and sample size per group. 113

TABLE B.1 : Maximum likelihood estimates from four Weibull mixture cure models for disease-specific survival. 136
TABLE B.2 : Results from the likelihood ratio test for nested models. 137
TABLE B.3 : Estimates of 5-year time-conditional survival probability from four Weibull mixture cure models for disease-specific survival adjusting for fixed tumor thickness (3.58mm) and varying ulceration status. 138

TABLE C.1 : An excerpt of the data from a randomized, placebo-controlled study on 59 epileptic patients with seizure counts measured every 2 weeks over an 8 week period (Thall and Vail, 1990). 172

LIST OF ILLUSTRATIONS

FIGURE 2.1 : Expected type I error of 5% with the 95% confidence interval for 10,000 datasets and estimated type I error for the GM model test statistic (and 95% confidence intervals) for no censoring, 10%, and 35% uniform random censoring. 31
FIGURE 2.2 : Estimated power for the LM and QM test statistics with and without censoring. 32
FIGURE 2.3 : Estimated power for the GM model test statistic. 33
FIGURE 2.4 : Nonparametric Kaplan-Meier estimated 3-year log time-conditional survival probabilities given 0, 6, 12, and 18 months after diagnosis for Stage II patients. 37
FIGURE 2.5 : Nonparametric Kaplan-Meier estimated 3-year log time-conditional survival probabilities given 0, 6, 12, and 18 months after diagnosis for patients by procedure: (1) Pathologically staged (some nodal procedure) and (2) Clinically staged (no nodal procedure). 39
FIGURE 2.6 : Nonparametric Kaplan-Meier estimated 3-year log time-conditional survival probabilities given 0, 6, 12, and 18 months after diagnosis for patients by ulceration status: (1) Not ulcerated and (2) Ulcerated. 39
FIGURE 2.7 : Nonparametric Kaplan-Meier estimated 3-year log time-conditional survival probabilities given 0, 6, 12, and 18 months after diagnosis for patients in one of four groups: (1) Pathologically staged (some procedure) and not ulcerated, (2) Pathologically staged (some procedure) and ulcerated, (3) Clinically staged (no nodal procedure) and not ulcerated, and (4) Clinically staged (no nodal procedure) and ulcerated. 43

FIGURE 3.1 : The unadjusted survival estimate from the parametric Weibull distribution and the empirical Kaplan-Meier survival function for the SEER esophageal sample. 70
FIGURE 3.2 : Estimated 5-year time-conditional survival probability given increasing time survived for mean tumor length from the Weibull distribution based on the SEER esophageal sample. 71
FIGURE 3.3 : Estimated 5-year time-conditional survival probability, given that survival is greater than 1, 2, and 3 years after diagnosis, for increasing tumor length from the Weibull distribution based on the SEER esophageal sample. . . 72
FIGURE 3.4 : Estimated 5-year time-conditional survival probability given increasing time survived for tumor length at 2, 5, 10, and 15 cm from the Weibull distribution based on the SEER esophageal sample. 73
FIGURE 3.5 : Estimated time-conditional survival probability, given that survival is greater than 2 years after diagnosis, as a function of Δ evaluated at mean tumor length from the Weibull distribution based on the SEER esophageal sample. 74
FIGURE 3.6 : Estimated 2-, 3-, 4-, and 5-year time-conditional survival probability, given that survival is greater than 2 years after diagnosis, for increasing tumor length from the Weibull distribution based on the SEER esophageal sample. 75
FIGURE 3.7 : Empirical survivor function for the SEER melanoma sample. 77
FIGURE 3.8 : Estimated 5-year time-conditional survival for the SEER melanoma sample with four cases based on the logistic-Weibull cure model. 81

CHAPTER 1

INTRODUCTION

This dissertation develops statistical methods for time-conditional survival probability and for equally spaced count data. In Chapter 2, we derive the asymptotic distribution of a vector of nonparametric estimators and use weighted least squares methodology for the analysis of time-conditional survival probabilities. In Chapter 3, we derive the asymptotic distribution of time-conditional survival probability estimators from a Weibull parametric regression model and from a Logistic-Weibull cure model, adjusting for continuous covariates. We implement the weighted least squares methodology to assess relevant hypotheses. In Chapter 4, we implement a maximum-likelihood based method for the analysis of equally spaced longitudinal count data with over-dispersion.

1.1. Time-Conditional Survival Probability Methods

This work was motivated by increased attention in the medical literature on conditional survival. We distinguish between two types of conditional survival probabilities. The first refers to those probabilities that condition on fixed covariates at time of diagnosis (e.g., Xu and O'Quigley, 2000). The second, which we refer to as time-conditional survival probabilities, condition on time survived and will be the focus of the work here. With earlier detection, better therapies for diseases, and more systematic tracking, patients in recent years have been surviving longer and information on their long-term follow-up is more readily available. With patients living longer, there is interest in estimating the probability of survival not from a patient's time of diagnosis, but rather from her/his present state sometime after diagnosis. Time-conditional survival probability is defined as the probability of surviving at least an additional Δ years given that a patient has already survived a years. As described in further detail in Chapter 2, this probability can be estimated by the ratio of the a- and $(\Delta + a)$-year estimated survival probabilities from a single Kaplan-Meier survivor function (Kaplan and Meier, 1958).

We create a statistical framework for investigating time-conditional survival probability by developing additional methodological approaches to address the relationship between estimated time-conditional survival probabilities, time survived, and patient prognostic factors. While the work presented here focuses on applications in oncology, it can be applied to time-to-event data in other

disciplines.

1.1.1. Nonparametric Methods

Time-to-event data generally contain observations that are censored. Censoring occurs in situations when a patient has not yet experienced an event and is known to be alive up to a particular time. When all that is known is that a patient is alive at a given point in time, that patient's survival data is right censored. In this work, we use the survival function, defined as the probability of surviving beyond a specified time, to estimate time-conditional survival probability. For the nonparametric approach, the Kaplan-Meier Product-Limit methodology is used to estimate the survivor function, which is then used to estimate survival probabilities (Kaplan and Meier, 1958).

Time-conditional survival probabilities are an alternative measure of future survival by accounting for time elapsed from diagnosis and are estimated as a ratio of survival probabilities. Relative survival is also defined as a ratio of probabilities in a target population relative to the expected survival probability in a comparable general population over a given follow-up period (Dickman and Adami, 2006). The estimate of five-year relative survival is the ratio of the estimated five-year survival probability for the target population divided by the expected five-year survival probability in the general population that is assumed to be fixed (e.g., Ederer, Axtell, and Cutler, 1961; Hakulinen, 1982). Dickman et al., 2004 describe four approaches to estimate a regression model for relative survival using maximum likelihood methodology. In our work, we use the weighted least squares methodology to analyze nonparametric and parametric based estimators of time-conditional survival probabilities.

Over the last two decades, clinical investigators have presented point estimates and corresponding 95% confidence limits for time-conditional survival probabilities. Clinical investigators report that patients who have survived for some time beyond diagnosis are more interested in estimates of time-conditional survival probabilities because these estimates offer more relevant prognostic information than estimates from traditional survival probabilities computed using time from initial diagnosis (e.g., Xing et al., 2010). Clinical research publications increasingly present estimates of time-conditional survival probabilities. A topic search for "conditional survival" on Web of Science conducted in early 2015 revealed approximately 150 articles published in the past five years (2010-2014) compared to approximately 50 articles published in 2005–2009 and less than 30 ar-

ticles published in 2000–2004. The interest in these estimates demonstrates its strong relevance to many clinical settings and highlights the importance of developing modeling methodology for time-conditional survival probabilities.

In Chapter 2, we develop the asymptotic distribution for estimates of log time-conditional survival probabilities. The asymptotic distribution facilitates the extension of statistical tools from estimation to the statistical testing of different hypotheses of interest. We base our methods for hypothesis testing and model fitting on the work of Grizzle, Starmer, and Koch, 1969, who used weighted least squares as part of a regression modeling strategy for proportions and proposed test statistics for evaluating simplified models. Koch, Johnson, and Tolley, 1972 applied this approach to survival probabilities. With modifications, we apply their approach to time-conditional survival probabilities. We use weighted least squares to develop a test statistics for relevant hypotheses, e.g. a multivariate omnibus test of pairwise differences. We show that the proposed test statistics for evaluating the relationship between time-conditional survival probabilities and additional time survived have central χ^2-distributions under the null hypotheses. Further, we conducted simulation studies to assess the empirical probability of making a type I error for one of the hypotheses tests developed and to assess the power of the various models and statistics proposed.

Additionally, we used weighted least squares techniques to fit regression models for the log time-conditional survival probabilities as a function of time survived after diagnosis to address clinically relevant questions. Quadratic, linear, and global mean models are used to explore the relationship between log time-conditional survival probabilities and time survived. To include discrete, categorical covariates, we develop a parametric framework for multivariable models. To avoid the problem of multiple testing due to comparisons among covariate patterns resulting from either categorical variables or categorization of continuous variables (Bennette and Vickers, 2012), we propose an overall test of significance in addition to pairwise comparisons. Population based survival data from patients with melanoma are used to illustrate the proposed methodology by evaluating survival in patients who underwent clinical staging versus pathological staging (Balch et al., 2001).

In contrast, consider an alternative approach for nonparametric inference using median residual lifetimes with censoring proposed by Jeong, Jung, and Costantino, 2008. As with time-conditional survival probabilities, they draw the comparison between information at diagnosis and at a time after diagnosis. For example, consider a patient's interest in their estimate of expected survival

and the impact of first-line treatment on life expectancy from *diagnosis* in contrast with a patient's interest in their estimate of residual life expectancy sometime *after diagnosis* and the impact of additional, second-line treatment.

To obtain an estimate of median residual lifetime for censored survival data, the authors' first model the survivor function. Then their approach infers the median of remaining lifetimes among survivors beyond time t at a fixed time point, t_0, to obtain an estimate of the median residual life function evaluated at t_0. To compute the median residual lifetime, the authors compute the residual lifetime for a patient who has survived beyond t_0 as $S(t \mid t_0) = S(t + t_0)/S(t_0)$ for $t_0 \geq 0$. As given in their manuscript by Equation 2, they then obtain the estimated median of the residual lifetime distribution at t_0 by solving the equation $\hat{u}(\theta_{t_0}) = 0$ for θ_{t_0} where

$$\hat{u}(\theta_{t_0}) = \hat{S}(t_0 - \theta_{t_0}) - 0.5\hat{S}(t_0),$$

and where $\hat{S}(t)$ is the Kaplan-Meier estimator of $S(t)$ (Jeong, Jung, and Costantino, 2008).

The resulting estimate provided to patients and physicians is the median residual life in years, which is more intuitive for patients to understand than a survival probability. However, a limitation of this approach is the influence of the proportion of censored observations as the median failure time cannot be theoretically defined until the minimum of the survival curve reaches 0.5 (Jeong, Jung, and Costantino, 2008). This issue is addressed in the later work by Park, Jeong, and Lee, 2012. As will be demonstrated in the application of Chapter 2, time-conditional survival probability under the nonparametric Kaplan-Meier framework is advantageous as it can be estimated irrespective of the minimum of the survival curve.

Further, in their paper, Jeong, Jung, and Costantino, 2008 note that the methods of comparing median residual life functions over the entire follow-up period that they developed did not address issues of multiple comparisons. In our work, we developed an omnibus test of contrasts and other hypotheses tests along with providing estimators of the covariance matrix to allow researchers to address issues of multiple comparisons. Lastly, the authors note the need for future research to develop a regression model that would take into account continuous prognostic factors and develop such an approach using regression on quantile residual life (Jung, Jeong, and Bandos, 2009). This is similar to the development of time-conditional survival where we began with the nonparametric

Kaplan-Meier framework in Chapter 2 and extended the approach to allow for continuous covariates in the parametric framework in Chapter 3.

1.1.2. Parametric Survival Methods

Parametric models are used by researchers for time-to-event data in the estimation of model parameters and related functions such as the parametric hazard function. Modeling survival time without the inclusion of covariates provides an estimate of the survival experience on the assumption that the underlying population is homogeneous. Incorporating covariates allows for the study of heterogeneous populations that may characterize observational studies based in disease registries, rather than populations from clinical trials with strict inclusion and exclusion criteria. Covariates can be incorporated by modeling the natural logarithm of survival time (e.g., Weibull or log logistic regression models) or using an accelerated failure-time model. In both cases, when the parametric model provides a good fit for the data, the estimates from the model are often more precise than those from the nonparametric setting because they are based on fewer parameters (e.g., Lambert and Royston, 2009).

In Chapter 2, we develop the large sample distribution for log time-conditional survival and pairwise tests for differences among a set of estimates. We also use weighted least squares to model a profile of estimates as a function of time survived and compare time-conditional survival estimates across groups of patients. Stratifying patients into groups based on covariate patterns requires using categorical variables or categorizing continuous variables. For example, Barchielli et al., 1994 categorized age at time of diagnosis into 5-year groups to evaluate the prognostic effect of this variable as opposed to evaluating it as a continuous covariate. Stratification assumes that patients falling into one stratum are homogeneous and, therefore, have a homogeneous risk for the outcome (Bennette and Vickers, 2012). However, ignoring variability within the stratum leads to a loss of information and reduces the power of a test of association (Greenland, 1995). Therefore, we extend the work in Chapter 2 by developing methods under parametric assumptions.

Chapter 3 develops methods for parametric time-conditional survival probability. It extends the methodology of Chapter 2 by allowing for the inclusion of multiple covariates, including continuous variables, in a single regression model for time-conditional survival probability. Two regression models are considered. In the first model, covariate adjusted time-conditional survival estimation

is based on the log-linear model for the relationship between time survived from diagnosis and covariates of interest. It is important to note that a survival model that tends to zero with increasing time after diagnosis will be appropriate for a disease with a generally poor prognosis. If, on the other hand, long-term survival is of interest due to patients surviving longer and having improved prognoses, a model that allows for a non-zero probability of indefinite survival or cure will better fit the data. The second regression is a parametric cure model where the underlying population is a mixture of patients who experience the event of interest and those who do not. The Weibull distribution is used to illustrate the methods which are applied to esophageal cancer data (Weibull regression model with covariates) and to melanoma data (Logistic-Weibull cure model).

A recent paper by Hieke et al., 2015 emphasized the usefulness of the conditional survival concept to provide information on the evolution of prognosis over time. In their application, these authors used Kaplan-Meier survival estimation to analyze data from multiple myeloma patients stratified by age groups and disease stage. Further, while they stated that methods to estimate conditional survival exist using Kaplan-Meier and Cox regression, they did not note uses of parametric regression models in the estimation of conditional survival. This indicates that a methodological gap remains in nonparametric estimation adjusting for continuous covariates which can be addressed using a parametric statistical approach.

When assessing hypothesis testing based on conditional survival methods, Hieke et al., 2015 show estimates in Figure 2B with a 5-year conditional survival profile plotted for each age strata and 95% confidence intervals around each point estimate of time since diagnosis in years. In our work, we develop nonparametric and parametric methods that account for the correlation among time-conditional survival probabilities through the hypothesis testing framework by incorporating the covariance matrix into the test statistic.

Parast, Cheng, and Cai, 2011, 2012 have developed methodology for incorporating short-term outcome information to predict long-term disease outcomes where the long-term event of interest is time to a terminal event such as death and the short-term event is time to a non-terminal event. These authors propose methods for incorporating censored short-term event information to predict long-term survival beyond the parametric models in a multi-state survival setting, which may lead to invalid prediction if the model assumptions do not hold. In their earlier work (2011) they developed nonparametric methods to predict the long-term outcome given the short-term outcome

and information on a discrete marker. However, similar to the limitation in the time-conditional survival methods developed in Chapter 2, these methods cannot account for information from one or more continuous covariates. The authors then extended work by others and proposed a flexible approach that allows for the inclusion of longitudinal predictor information collected such as repeated biomarker measurements (2012). As the authors note, including information about a short-term outcome in addition to genetic or biomarker measurements may lead to an improved ability to predict long-term survival. Accordingly, future research on implementing competing risks models in the estimation of time-conditional survival needs to be explored.

1.2. Longitudinal Count Data Methods

Longitudinal count data are often encountered in scientific studies. Common features of longitudinal count data include intra-subject correlation and over-dispersion. Intra-subject correlation is due to similarities between the repeated measurements on each participant. When the variance is larger than expected for the assumed distribution of the outcome variable then over-dispersion is observed (Efron, 1992).

As noted by Farewell and Farewell, 2012, one approach to modeling longitudinal Poisson count data is using a generalized linear mixed model with Poisson distributions conditional on random effects. They note that a marginal modeling approach may be preferred for cases where the effect of explanatory variables at the population-averaged level (marginal effects) is of interest as opposed to subject-specific effects. Heagerty and Kurland, 2001 showed that marginal modeling is more robust for the estimation of regression parameters as compared to subject-specific covariate effects when there is a departure from the underlying random effects structure.

Marginal modeling can be implemented using generalized estimating equation (GEE) methods (Solis-Trapala and Farewell, 2005). Two considerations when using the GEE approach are that GEE methods do not give the researcher an understanding of the sources of variation and, unlike parametric maximum likelihood estimation, there is reduced efficiency (Farewell and Farewell, 2012). Given these limitations, Farewell and Farewell, 2012 developed methods to analyze such data using the Dirichlet negative multinomial distribution. From their simulation study to evaluate the model robustness and finite-sample behavior, the authors found that the Dirichlet negative multinomial regression was preferred over the GEE method. When applying this methodology to their

data, Farewell and Farewell, 2012 also fit a Poisson generalized linear mixed model and found that the estimated coefficients and standard errors were similar to their Dirichlet negative multinomial model. Our approach offers an alternative to their methods and to the GEE approach.

Over-dispersed count data are often encountered in longitudinal studies. This may be present in the context of the number of patients with epileptic seizures (Farewell and Farewell, 2012; Thall and Vail, 1990) or the number of patients who had transplants performed. Over-dispersion occurs when the variability is larger than the standard Poisson variability that is expected. However, few likelihoods are available for the simulation and analysis of such data (Efron, 1992). Therefore, we provide a maximum likelihood approach to model longitudinal Poisson count outcomes.

In Chapter 4, we develop an approach for maximum likelihood analysis of longitudinal discrete data with over-dispersion. We implement a likelihood proposed for simulation of over-dispersed random variables with specified marginal means and product correlations by Guerra and Shults, 2014. The key features of this approach are first-order antedependence and linearity of the conditional expectations. We also assume a Markovian model of first order, implying that the value of an outcome on a subject at a specific measurement occasion only depends on the value at the previous measurement occasion. Our maximum likelihood approach using the Poisson model for count data benefits from a simple interpretation of regression parameters, like that in GEE analysis of count data.

As described elsewhere (e.g., Shults et al., 2006 Appendix A), the maximum likelihood approach for count data requires information on the correlation between adjacent measurements on each subject. While Guerra and Shults, 2014 developed general simulation methods allowing for different patterns of correlation, we focus our maximum likelihood approach to Poisson count data with a first-order autoregressive (AR(1)) correlation structure. Given the specified marginal means and adjacent correlations, the AR(1) correlation structure is induced and the marginal distributions are over-dispersed relative to the Poisson distributions. Under the AR(1) structure, it is assumed that the adjacent intra-subject correlations are constant. This assumption is appropriate when it is reasonable to assume that two count outcomes that are measured closer in time will be more highly correlated, because they are assumed to be more similar, rather than if they are farther apart in time, in which case they are assumed to be less similar.

We obtain likelihood-based estimating equations for the regression and correlation parameters. Simulations are conducted to demonstrate that the approach has good statistical properties. This approach is applied to the analysis of health policy data on doctor visits (StataCorp LP, 2013; Winkelmann, 2004) and to seizure data (Farewell and Farewell, 2012; Thall and Vail, 1990).

1.3. Dissertation Structure

This dissertation is structured as follows. Chapter 2 defines the framework for nonparametric methods to assess time-conditional survival probability. This includes the development of the asymptotic distribution for a vector or profile of time-conditional survival probabilities, a flexible framework for hypothesis testing using point estimates, and regression modeling to address clinically relevant questions. Chapter 3 builds a parametric framework for time-conditional probability and incorporates multiple discrete and continuous covariates in modeling time-conditional survival profiles. These methods are applied using a Weibull regression model for data where survival tends to zero with increasing time after diagnosis and a Logistic-Weibull cure model for data where there is evidence of cure in a fraction of patients. Chapter 4 discusses a new approach for maximum likelihood-based analysis of correlated count data with over-dispersion. This maximum likelihood approach assesses the problem of over-dispersion in Poisson data with an AR(1) structure. Such longitudinal outcomes can be found in medical research in situations where measurements on a subject are captured at pre-specified occasions over time. Poisson regression is often used for analysis of count data, but would not be appropriate in an analysis of data characterized by over-dispersion. Key assumptions of the maximum likelihood approach include the first-order Markov property and the linearity of the conditional expectations for the conditional distributions. The proposed approach is applied in analysis of data on doctor visits and epilepsy seizure data (R code is available in the Appendix). Finally, Chapter 5 presents conclusions and envisions future work.

CHAPTER 2

NONPARAMETRIC TIME-CONDITIONAL SURVIVAL PROBABILITY

2.1. Introduction

Commonly reported statistics for cancer patients include estimates of survival and conditional survival probabilities. Over the past two decades, clinical investigators have also reported estimated time-conditional survival probabilities for patients who have already survived for a specified amount of time after the diagnosis or therapy of their disease. While conditional survival probabilities condition on covariates that were measured at the time of diagnosis (e.g., Xu and O'Quigley, 2000), time-conditional survival probabilities condition on the time already survived after diagnosis. Specifically, time-conditional survival probability is defined as the probability of surviving at least an additional x years given survival a years after diagnosis.

Typically, what is reported in the medical literature are the point estimates of the time-conditional survival probabilities and their associated 95% confidence limits that are based on the estimated Kaplan-Meier survival function (e.g., Choi et al., 2008; Merrill, Henson, and Ries, 1998; Xing et al., 2010). These estimates can be used to answer questions such as "What is the expected probability that a patient will survive an additional 5 years, given that she has already survived 5 years since diagnosis"? However, additional methods are needed to answer other clinically relevant questions of interest, such as "Does the expected probability that a patient will survive an additional 5 years *significantly* increase with increasing time post-diagnosis"? For example, is the expected probability of an additional 5 years survival significantly greater if the patient has survived 5 years than if she has only survived 1 year after diagnosis?

This type of question is of clinical interest because it has been observed (Ries et al., 2003) that for some cancers, the estimated time-conditional survival probabilities increase with an increasing number of years survived. More recently, Miller, Lynch, and Buckwalter, 2013 investigated 5-year conditional survival for a cohort of patients from the Surveillance, Epidemiology, and End Results (SEER) registry with osteosarcoma and Ewings sarcoma. They found that 5-year conditional survival was 74.8% at diagnosis and 91.4% given survival beyond 5 years after diagnosis. Wang et al., 2011a investigated 5-year conditional survival for a cohort of patients diagnosed with rectal cancer

from the SEER registry. For Stage I patients, the 5-year survival at diagnosis was 71% and 5-year conditional survival given survival beyond 5 years increased to 74%.

New methods are also needed to address questions about trends in the time-conditional survival probabilities over the course of a study. For patients who have survived some time after diagnosis, the probability of surviving an additional number of years may be different from an initial overall survival probability at diagnosis because the probability is not necessarily static (Choi et al., 2008). For example, is the change in probabilities linear, or quadratic? If we have different sub-groups of patients, does the change over time differ between the groups? Does the strength of this relationship differ for males versus females? In this chapter we develop methodology to answer such questions.

In Section 2, we define point estimates of time-conditional survival probabilities and their 95% confidence intervals, methods which have been published in the medical literature. In Section 3, we develop the asymptotic distribution of a vector of estimators of time-conditional survival probabilities using large sample distribution theory. We derive the estimate for the natural logarithm $(\log)$ of time-conditional survival probability and its estimated variance as a function of the number of years survived.

Grizzle, Starmer, and Koch, 1969 presented a general approach for the analysis of categorical data using linear models with weighted regression. This approach allowed for simplification in model formulation and hypothesis testing within the linear models framework. Koch, Johnson, and Tolley, 1972 applied linear regression models and weighted least squares methodology to the analysis of survival rates. In Section 4, we use Koch's regression modeling strategy using weighted least squares to analyze time-conditional survival probabilities. In particular, we develop Wald test statistics (Wald, 1943) to evaluate trends in time-conditional survival estimators that are relevant to patients, clinicians, and researchers.

In Section 5, we present results from simulations that assess power and the empirical probability of making a type I error for particular tests. We then apply the proposed methodology to a cancer study in Section 6, where we estimate time-conditional probabilities for melanoma patients as a function of time survived. Additionally, we evaluate the differences in time-conditional survival between groups of patients. In Section 7, we present some discussion and concluding remarks.

2.2. Estimation of Time-Conditional Survival Probabilities

2.2.1. Notation

Let n be the fixed number of individuals in a study. Define $T_i = \min(X_i, C_i)$, where X_i and C_i are the event and censoring times for the ith subject, respectively. We assume the censoring time, C_i, is independent of the event time, X_i. We observe the pair $(t_i, \delta_i), i = 1, \ldots, n$, where t_i is the time on study and $\delta_i = I(X_i \leq C_i)$ is the indicator variable for whether t_i is an event or a censoring time. Define J distinct event times to be $t_{(1)} < \cdots < t_{(J)}$ allowing for possible ties in the data. For each observed event time $t_{(j)}$ in the set of ordered event times, define n_j to be the number of subjects at risk at time $t_{(j)}$ and let d_j be the number of events observed at time $t_{(j)}$ among the n_j subjects at risk. To incorporate information on censoring, let w_j denote the number of observations that are (right) censored at times after the jth event time, but prior to the $(j+1)$th time.

2.2.2. Current approach

Assume we have non-informative censoring, where knowledge of an individual's censoring time provides no further information about the patient's likelihood of survival at a future time had they continued on the study. Under this assumption of non-informative censoring, the likelihood is given by

$$L(\pi_1, \ldots, \pi_N) \propto \prod_{j=1}^{J} \pi_j^{d_j} S(t_{(j-1)})^{d_j} S(t_{(j)})^{w_j},$$

where $\pi_j = \lim_{\Delta t \downarrow 0} P(t_{(j)}^- < T \leq t_{(j)}^- + \Delta t \mid T > t_{(j)}^-)$ is the conditional probability of an event at $t_{(j)}$, $j = 1, \ldots, J$, and where the survival function is given by $S(t) = P(T \geq t)$ such that $t_{(0)} = 0$ and $S(t_{(0)}) = 1$ (Lachin, 2000).

Survival beyond time $t_{(j)}$ requires a subject to be event-free beyond time $t_{(j-1)}$ and all previous times. If the survivor function is rewritten in terms of π_j, the likelihood can be simplified in the following product binomial form

$$L(\pi_1, \ldots, \pi_J) \propto \prod_{j=1}^{J} \pi_j^{d_j} (1 - \pi_j)^{n_j - d_j}.$$

The maximum likelihood estimator (MLE) of π_j is given by $\hat{\pi}_j = \frac{d_j}{n_j}$, $j = 1, \ldots, J$, and any two

estimators from the same sample, $\hat{\pi}_j$ and $\hat{\pi}_k$, where $1 \leq j < k \leq J$, are uncorrelated (Lachin, 2000).

Define the x given a time-conditional survival probability as the probability of surviving at least an additional x years given survival a years after diagnosis as

$$P\left(T > a + x \mid T > a\right) = \frac{P\left(T > a + x\right)}{P\left(T > a\right)} = \frac{S(a+x)}{S(a)}, \tag{2.1}$$

where $a \geq 0$ and $x > 0$. To provide shorthand notation for the time-conditional survival probability, let

$$CS\left(a + x \mid a\right) = P\left(T > a + x \mid T > a\right).$$

The times a and b = $a + x$ are not necessarily observed event times, however, both a and b should be chosen so that $0 \leq a, b \leq t_{(J)}$ and $a < b$. The time-conditional survival probability is estimated using the maximum likelihood estimators of the conditional probabilities, $\hat{\pi}_j = \frac{d_j}{n_j}$, $j = 1, \ldots, J$, by

$$\widehat{CS}(b \mid a) = \frac{\hat{S}(b)}{\hat{S}(a)} = \frac{\prod_{j:t_{(j)} \leq b}\left(1 - \hat{\pi}_j\right)}{\prod_{j:t_{(j)} \leq a}\left(1 - \hat{\pi}_j\right)} = \prod_{j:a < t_{(j)} \leq b}\left(1 - \hat{\pi}_j\right) = \prod_{j:a < t_{(j)} \leq b}\left(1 - \frac{d_j}{n_j}\right). \tag{2.2}$$

To derive the variance of the time-conditional survival probability, we use computations similar to those used to obtain Greenwood's formula (Greenwood, 1926) from the estimated Kaplan-Meier survivor function. The estimated variance is given by

$$\widehat{Var}\left(\widehat{CS}(b \mid a)\right) = \widehat{Var}\left(\frac{\hat{S}(b)}{\hat{S}(a)}\right) = \left(\widehat{CS}(b \mid a)\right)^2 \sum_{j:a < t_{(j)} \leq b} \frac{d_j}{n_j(n_j - d_j)}. \tag{2.3}$$

See Appendix Section A.1 for the derivation.

2.3. Distribution Theory

Clinical studies of time-conditional survival have used a profile of estimated time-conditional survival probabilities (shown here as a $p \times 1$ vector) given by

$$\widehat{\mathbf{CS}} = \left(\widehat{CS}_1(b_1 \mid a_1), \widehat{CS}_2(b_2 \mid a_2), \ldots, \widehat{CS}_p(b_p \mid a_p)\right)^T.$$

When $b_j = a_j + x$ for $j = 1, \ldots, p$, these estimators represent consecutively estimated x-year time-conditional survival probabilities. For example, 5-year time-conditional survival probabilities given survival beyond 1, 2, and 3 years after diagnosis are consecutive estimators that can be expressed as

$$\widehat{\mathbf{CS}}_{3\times 1} = \left(\widehat{CS}_1(6 \mid 1), \widehat{CS}_2(7 \mid 2), \widehat{CS}_3(8 \mid 3)\right)^T.$$

2.3.1. The choice of p

The choice of p is limited by the amount of follow-up data available, the timing of events, and the researcher's choice of x and a (refer to Equation 2.1). Consider a data set where 10 years of annual follow-up data is available. When researchers are interested in 5-year time-conditional survival probabilities, at most five distinct time-conditional survival estimates can be computed using a one year increment post baseline (time=0). Specifically, we assume that distinct estimates of survival are available for $S(1), \ldots, S(10)$, which allows for subsequent estimation of 5-year time-conditional survival probabilities given survival from 0 through 5 years after diagnosis.

For this example, the profile of distinct time-conditional survival estimates is given by

$$\widehat{\mathbf{CS}}_{6\times 1} = \left(\widehat{CS}_1(5 \mid 0), \widehat{CS}_2(6 \mid 1), \widehat{CS}_3(7 \mid 2), \widehat{CS}_4(8 \mid 3), \widehat{CS}_5(9 \mid 4), \widehat{CS}_6(10 \mid 5)\right)^T.$$

This profile represents estimates of 5-year time-conditional survival probabilities given survival at diagnosis (year 0) and beyond 1, 2, 3, 4, and 5 years after diagnosis. Note that $\widehat{CS}(5 \mid 0) = \hat{S}(5) = \hat{P}(T \geq 5)$. Therefore, the elements of the profile, and the covariance matrix that correspond to this term, will reflect that $\widehat{CS}(5 \mid 0)$ is a survival probability.

2.3.2. Asymptotic distributions

In this section we derive the asymptotic distribution of the natural logarithm (log) of the profile of estimators. Assume that n_j/n converges in probability to ω_j, $n_j/n \xrightarrow{p} \omega_j$. For fixed a and b, where $a < b$, the asymptotic distribution of the log of the estimated time-conditional survival probability, $\log \widehat{CS}(b \mid a)$, is given by

$$\sqrt{n}\frac{\left(\sum_{j:a<t_{(j)}\leq b} \log(1-\hat{\pi}_j) - \sum_{j:a<t_{(j)}\leq b} \log(1-\pi_j)\right)}{\sqrt{\sum_{j:a<t_{(j)}\leq b} \frac{\hat{\pi}_j}{\hat{\omega}_j(1-\hat{\pi}_j)}}} \xrightarrow{d} N(0,1), \tag{2.4}$$

where $n_j/n \xrightarrow{p} \omega_j$ as $n \to \infty$. This result follows from the consistency, asymptotic normality, and invariance properties of the maximum likelihood estimator of π_j. See Appendix Section A.3 for details.

Using the δ-method, the large sample expectation of the individual log time-conditional survival estimator is given by

$$E\left(\log \widehat{CS}(b \mid a)\right) \cong \sum_{j:a<t_{(j)}\leq b} \log\left(1-\pi_j\right),$$

and the large sample variance is given by

$$Var\left(\log \widehat{CS}(b \mid a)\right) \cong \sum_{j:a<t_{(j)}\leq b} \frac{\pi_j}{n_j(1-\pi_j)}.$$

Substituting the maximum likelihood estimator, $\hat{\pi}_j$, for π_j, the estimated mean and variance are then given by

$$\hat{E}\left(\log \widehat{CS}(b \mid a)\right) = \sum_{j:a<t_{(j)}\leq b} \log\left(1-\frac{d_j}{n_j}\right), \tag{2.5}$$

and

$$\widehat{Var}\left(\log \widehat{CS}(b \mid a)\right) = \sum_{j:a<t_{(j)}\leq b} \frac{d_j}{n_j(n_j-d_j)}, \tag{2.6}$$

respectively.

Define the general p-vector profile of log time-conditional survival probability estimators as

$$\log \widehat{\mathbf{CS}} = \left(\log \widehat{CS}_1(b_1 \mid a_1), \log \widehat{CS}_2(b_2 \mid a_2), \ldots, \log \widehat{CS}_p(b_p \mid a_p)\right)^T. \tag{2.7}$$

This profile is defined by a fixed difference between times b_i and a_i, such that $b_i - a_i = c$ for $i = 1, \ldots, p$. The asymptotic distribution of the estimator of the p-vector profile is then given by

$$\log \widehat{\mathbf{CS}} \xrightarrow{d} N\left(E(\log \widehat{\mathbf{CS}}), Var(\log \widehat{\mathbf{CS}})\right)$$

such that

$$E\left(\log \widehat{\mathbf{CS}}\right) \cong \log \mathbf{CS},$$

$$Var\left(\log \widehat{\mathbf{CS}}\right) \cong \mathbf{\Sigma}.$$

The formula for the estimate of each log time-conditional survival probability is given in Equation 2.5. See Appendix Section A.3 for more details regarding the derivation. Note that the inclusion of the survival probability at diagnosis, such as $\log \widehat{CS}_1(5 \mid 0)$, will need to be reflected as a survival probability in the estimated p-vector profile and in the estimation of the variance and covariance terms described below.

To describe the terms of the covariance matrix, $\mathbf{\Sigma}$, and the estimated covariance matrix, $\hat{\mathbf{\Sigma}}$, we define any two estimators of log time-conditional survival probabilities. These are given by

$$\log \widehat{CS}_l(b_l \mid a_l) = \log\left(\frac{\hat{S}(b_l)}{\hat{S}(a_l)}\right), \tag{2.8}$$

and

$$\log \widehat{CS}_m(b_m \mid a_m) = \log\left(\frac{\hat{S}(b_m)}{\hat{S}(a_m)}\right), \tag{2.9}$$

where $1 \leq l, m \leq J$ and where a_l, b_l, a_m, b_m are fixed times such that $0 \leq a_l \leq a_m \leq b_l \leq b_m \leq t_{(J)}$. As shown in Appendix Section A.3, the elements of the covariance matrix, Σ, for $l = m$, are given by

$$\Sigma_{ll} = Var\left(\log \widehat{CS}_l(b_l \mid a_l)\right) = \sum_{j: a_l < t_{(j)} \leq b_l} \frac{\pi_j}{n_j(1-\pi_j)}$$

and for $l \neq m$ are given by

$$\Sigma_{lm} = Cov\left(\log \widehat{CS}_l(b_l \mid a_l), \log \widehat{CS}_m(b_m \mid a_m)\right) = \sum_{j: a_m < t_{(j)} \leq b_l} \frac{\pi_j}{n_j(1-\pi_j)}.$$

The covariance matrix, $\mathbf{\Sigma}$, is estimated by $\hat{\mathbf{\Sigma}}$ where the elements of $\hat{\mathbf{\Sigma}}$ for $l = m$ are given by

$$\hat{\Sigma}_{ll} = \widehat{Var}\left(\log \widehat{CS}_l(b_l \mid a_l)\right) = \sum_{j: a_l < t_{(j)} \leq b_l} \frac{d_j}{n_j(n_j - d_j)} \tag{2.10}$$

and for $l \neq m$ are given by

$$\hat{\Sigma}_{lm} = \widehat{Cov}\left(\log \widehat{CS}_l(b_l \mid a_l), \log \widehat{CS}_m(b_m \mid a_m)\right) = \sum_{j: a_m < t_{(j)} \leq b_l} \frac{d_j}{n_j(n_j - d_j)}. \tag{2.11}$$

See Appendix Section A.2 for a detailed derivation. The covariance is 0 if a_l, b_l, a_m, b_m are non-overlapping fixed times such that $0 \leq a_l < b_l < a_m < b_m \leq t_{(J)}$. Refer to Appendix Section A.3 for

more details on the large sample distribution.

2.4. Hypothesis Testing Using Weighted Least Squares

We use the weighted least squares methodology of Grizzle, Starmer, and Koch, 1969 and Koch, Johnson, and Tolley, 1972 to evaluate the relationship between time-conditional survival probabilities and additional time survived. We propose test statistics to assess these relationships that have central χ^2-distributions under the null hypotheses. For example, is there a linear relationship between the probabilities and additional time survived post-baseline?

We focus on four clinically relevant research questions using hypothesis testing. The first question when considering a single time-conditional survival probability profile is whether the profile is constant. The null hypothesis that the time-conditional survival probabilities in the profile are the same indicating that increasing survival time after diagnosis (a) does not result in either an increased (or decreased) likelihood of surviving an additional number of years. We first propose an omnibus test of contrasts for adjacent time-conditional survival probabilities.

When the null hypothesis is rejected, there could be a linear or quadratic relationship between the time conditional survival probabilities and additional time survived. The second objective is to further consider the shape of the time-conditional survival probability profile. We propose a series of hypothesis tests to identify the most parsimonious regression model to describe the relationship between the log time-conditional survival probabilities and time survived after diagnosis. Specifically, we develop three regression models to evaluate the relationship between log time-conditional survival probabilities and time survived: the quadratic model (QM), the linear model (LM), and the global mean (GM) model. A quadratic relationship would suggest that there is a greater benefit with increasing time survived post diagnosis than if the relationship were linear. Note that the GM model is equivalent to the constant profile described for the first hypothesis test.

The third objective is to assess whether time-conditional survival profiles for independent strata are the same and are constant. To do so, we estimate multiple time-conditional survival probability profiles, one for each strata defined by categorical covariates. Under the null hypothesis, the profiles are the same for all strata and are constant. This would indicate that there is no change in the likelihood of surviving additional years with increased time after diagnosis in each stratum.

The fourth and final objective is to apply a modeling strategy to evaluate differences in time-conditional survival probabilities based on the covariates used to create the strata. For example, to determine if observed linear relationships differ significantly between males and females, we develop a set of hypothesis tests based a multivariable model framework allowing for the inclusion of interaction terms. Under the null hypothesis of no interaction, if we have two independent binary covariates used to create four strata, for example, then the model is adequately represented by an additive model without the interaction term.

2.4.1. General Framework

In the case where there is a single population, the log time-conditional survival proportions are defined by a $p \times 1$ vector of estimated probabilities, $\log \widehat{\mathbf{CS}}$, which was defined in Equation 2.7. Assume a regression model under the null hypothesis,

$$E(\log \widehat{\mathbf{CS}}) = \mathbf{X}\boldsymbol{\beta},$$

where $\mathbf{X}$ ($p \times d$) is a design matrix of rank $d \leq p$ and $\boldsymbol{\beta}$ is the corresponding vector of parameters. This model can be fit using weighted least squares where the estimates of $\boldsymbol{\beta}$ are obtained by weighting the estimating equations for $\boldsymbol{\beta}$ by the inverse of the estimated $p \times p$ covariance matrix, $\hat{\boldsymbol{\Sigma}}$, defined in Section 2.3.2, so that

$$\hat{\boldsymbol{\beta}} = \left[\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\mathbf{X}\right]^{-1}\mathbf{X}'\hat{\boldsymbol{\Sigma}}^{-1}\log \widehat{\mathbf{CS}}.$$

To test for overall regression, the null hypothesis is given by

$$H_0 : \log \mathbf{CS} = \mathbf{X}\boldsymbol{\beta},$$

and the test statistic is given by

$$TS(\log \mathbf{CS} = \mathbf{X}\boldsymbol{\beta}) = \left(\log \widehat{\mathbf{CS}} - \mathbf{X}\hat{\boldsymbol{\beta}}\right)' \hat{\boldsymbol{\Sigma}}^{-1} \left(\log \widehat{\mathbf{CS}} - \mathbf{X}\hat{\boldsymbol{\beta}}\right).$$

For large n, the test statistic is approximately distributed as a central χ^2 with degrees of freedom $rank(\mathbf{X})$ under the null hypothesis. Under the alternative, the test statistic is distributed as a non-

central χ^2.

2.4.2. Contrasts of profile-based differences

A common objective in the medical literature is to evaluate differences among estimated time-conditional survival probabilities. We define the log time-conditional survival probability as

$$\log CS_i = \log CS_i(b_i \mid a_i) = log\left(\frac{S(b_i)}{S(a_i)}\right), \qquad i = 1, \ldots, p,$$

such that $a_1 < a_2 < \ldots < a_p < b_1 < b_2 < \ldots < b_p$. Then, define the null hypothesis where, with increasing time after diagnosis, the log time-conditional survival probabilities are all equal. Under this null hypothesis, we expect the vector of pairwise differences between adjacent log time-conditional survival probabilities to be zero. The null and alternative hypotheses are given by

$$H_0 : \begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{pmatrix} \begin{pmatrix} \log CS_1 \\ \log CS_2 \\ \vdots \\ \log CS_p \end{pmatrix} = \mathbf{X}_C \log \mathbf{CS} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{2.12}$$

and

$$H_1 : \begin{pmatrix} \log CS_1 - \log CS_2 \\ \vdots \\ \log CS_{p-1} - \log CS_p \end{pmatrix} = \begin{pmatrix} \Delta_1 \\ \vdots \\ \Delta_{p-1} \end{pmatrix} = \mathbf{\Delta}.$$

The $(p-1) \times 1$ vector of estimated pairwise differences is then given by

$$\hat{\mathbf{\Delta}} = \left(\log \widehat{CS}_1 - \log \widehat{CS}_2, \log \widehat{CS}_2 - \log \widehat{CS}_3, \ldots, \log \widehat{CS}_{p-1} - \log \widehat{CS}_p\right)',$$

where each of the differences in log time-conditional survival probabilities is estimated using the Kaplan-Meier survivor function estimator, $\hat{S}(t)$ (see Equation 2.2).

From the distribution of $\log \widehat{\mathbf{CS}}$, it follows that the asymptotic distribution of the vector of pairwise differences is given by

$$\mathbf{X}_C \log \widehat{\mathbf{CS}} \xrightarrow{d} N_{p-1}\left(\mathbf{X}_C \log \mathbf{CS}, \mathbf{X}_C \mathbf{\Sigma} \mathbf{X}_C'\right).$$

The weighted test statistic for the test of this null hypothesis is given by

$$TS(C) = \hat{\boldsymbol{\Delta}}' \left(\mathbf{X}_C \hat{\boldsymbol{\Sigma}} \mathbf{X}_C' \right)^{-1} \hat{\boldsymbol{\Delta}},$$

where $\mathbf{X}_C$ is the $(p-1) \times p$ matrix given in Equation 2.12. For large n, $TS(C)$ is approximately distributed as a central $\chi^2(p-1)$ under the null hypothesis and is distributed as a non-central χ^2 under the alternative hypothesis.

When we reject the above null hypothesis that all adjacent pairwise differences are zero, we can further evaluate each pairwise difference. Each null hypothesis is that there is no difference in the adjacent pairwise difference of time-conditional survival probability estimators given by

$$\Delta_h = \log CS_h - \log CS_{h+1} = 0, \qquad h = 1, \ldots, p-1.$$

The univariate χ^2 test statistic is given by

$$TS(C_h) = \frac{\left(\log \widehat{CS}_h - \log \widehat{CS}_{h+1} \right)^2}{\widehat{Var}\left(\log \widehat{CS}_h \right) + \widehat{Var}\left(\log \widehat{CS}_{h+1} \right)},$$

and under the null hypothesis, the test statistic is approximately distributed as a central $\chi^2(1)$. When multiple such independent pairwise tests are computed, we adjust the significance level using the Bonferroni correction to achieve a total Type I error probability no greater than 5%.

2.4.3. Regression models for time-conditional survival probabilities

To assess whether there is evidence of a relationship between log time-conditional survival probabilities and time survived after diagnosis, we develop three hypothesis tests. These models and hypothesis tests allow researchers to investigate whether the profile of probabilities follows a quadratic model (QM), a linear model (LM), or a global mean (GM) model. When performing these goodness-of-fit tests, we begin with fitting the QM. This approach is appropriate when there are at least four time-conditional survival probabilities ($p \geq 4$) ensuring adequate degrees of freedom to test the QM hypothesis.The goal is to find the most parsimonious model to fit the profile by subsequently testing the remaining models. Without loss of generality, we assume that time consistently increases by 1 unit to define the time-conditional survival probabilities and reflect this in the design matrix.

Quadratic Model. The estimated log time-conditional survival probabilities are first compared to QM where the null and alternative hypotheses are given by

$$H_0 : \log \mathbf{CS} = \mathbf{X}_Q \boldsymbol{\beta}_Q \qquad and \qquad H_1 : \log \mathbf{CS} = \boldsymbol{\theta}.$$

The parameter vector is $\boldsymbol{\beta}_Q = (\beta_0, \beta_1, \beta_2)'$ representing the intercept, linear, and quadratic parameters and the $p \times 3$ design matrix is given by

$$\mathbf{X}_Q = \begin{pmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ \vdots & \vdots & \vdots \\ 1 & p-1 & (p-1)^2 \end{pmatrix}.$$

Under the alternative hypothesis, the vector, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_p)'$, of log time-conditional survival probabilities is estimated from the Kaplan-Meier survivor function estimator, $\hat{S}(t)$ (see Equation 2.2).

The weighted least squares estimate of $\boldsymbol{\beta}_Q$ is given by

$$\hat{\boldsymbol{\beta}}_Q = \left[\mathbf{X}_Q' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_Q\right]^{-1} \mathbf{X}_Q' \hat{\boldsymbol{\Sigma}}^{-1} \log \widehat{\mathbf{CS}},$$

where $\hat{\boldsymbol{\Sigma}}$ is the estimated covariance matrix of $\log \widehat{\mathbf{CS}}$ under the alternative hypothesis (Koch, Johnson, and Tolley, 1972). The test statistic is given by

$$TS(Q) = \left(\log \widehat{\mathbf{CS}} - \mathbf{X}_Q \hat{\boldsymbol{\beta}}_Q\right)' \hat{\boldsymbol{\Sigma}}^{-1} \left(\log \widehat{\mathbf{CS}} - \mathbf{X}_Q \hat{\boldsymbol{\beta}}_Q\right),$$

where the design matrix, $\mathbf{X}_Q$, is $p \times 3$. For large samples, $TS(Q)$ is approximately distributed as a central $\chi^2(p-3)$ under the null hypothesis. Under the alternative, $TS(Q)$ is distributed as a non-central χ^2. When the null hypothesis is not rejected, we conclude that the profile of log time-conditional survival probabilities does not significantly differ from the quadratic model.

Linear Model. To assess whether the log time-conditional survival probabilities are a linear function of the number of years survived, the null and alternative hypotheses as given by

$$H_0 : \log \mathbf{CS} = \mathbf{X}_L \boldsymbol{\beta}_L \qquad \text{and} \qquad H_1 : \log \mathbf{CS} = \boldsymbol{\theta}.$$

Similar to the QM, the weighted least squares estimate of $\boldsymbol{\beta}_L = (\beta_0, \beta_1)'$, the intercept and linear parameters, is given by

$$\hat{\boldsymbol{\beta}}_L = \left[\mathbf{X}_L' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_L\right]^{-1} \mathbf{X}_L' \hat{\boldsymbol{\Sigma}}^{-1} \log \widehat{\mathbf{CS}},$$

where $\hat{\boldsymbol{\Sigma}}$ is the estimated covariance matrix of $\log \widehat{\mathbf{CS}}$ and the $p \times 2$ design matrix is given by

$$\mathbf{X}_L = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ \vdots & \vdots \\ 1 & p-1 \end{pmatrix}.$$

For this hypothesis test,

$$TS(L) = \left(\log \widehat{\mathbf{CS}} - \mathbf{X}_L \hat{\boldsymbol{\beta}}_L\right)' \hat{\boldsymbol{\Sigma}}^{-1} \left(\log \widehat{\mathbf{CS}} - \mathbf{X}_L \hat{\boldsymbol{\beta}}_L\right)$$

is the test statistic where the design matrix is $p \times 2$. The test statistic, $TS(L)$, is approximately distributed as a central $\chi^2(p-2)$ under the null hypothesis. If $TS(Q)$ does not lead to rejection of the QM hypothesis but $TS(L)$ does lead to rejection of the LM hypothesis, then the quadratic model is the more parsimonious model to fit the data. Alternatively, when the null hypothesis is not rejected using $TS(L)$, we conclude that the profile of log time-conditional survival probabilities does not significantly differ from the linear model.

Global Mean. Under the GM hypothesis, all time-conditional survival probabilities are equal to the same parameter. Further, the profile is adequately represented by a line with slope 0. The hypotheses are given by

$$H_0 : \log \mathbf{CS} = \mathbf{X}_G \beta_G \qquad \text{and} \qquad H_1 : \log \mathbf{CS} = \boldsymbol{\theta},$$

and the parameter β_G is estimated by $\hat{\beta}_G$ where

$$\hat{\beta}_G = \left[\mathbf{X}_G' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_G\right]^{-1} \mathbf{X}_G' \hat{\boldsymbol{\Sigma}}^{-1} \log \widehat{\mathbf{CS}},$$

and the $p \times 1$ design matrix, $\mathbf{X}_G$, is a vector of ones. Then,

$$TS(G) = \left(\log \widehat{\mathbf{CS}} - \mathbf{X}_G \hat{\beta}_G\right)' \hat{\mathbf{\Sigma}}^{-1} \left(\log \widehat{\mathbf{CS}} - \mathbf{X}_G \hat{\beta}_G\right),$$

is the weighted multiple regression test statistic where $\mathbf{X}_G$ is a $p \times 1$ vector under the GM model. For large samples, $TS(G)$ is approximately distributed as a central $\chi^2(p-1)$ under the null hypothesis and is distributed as a non-central χ^2 under the alternative hypothesis. If $TS(L)$ does not lead to rejection of the LM hypothesis but $TS(G)$ does lead to rejection of the GM hypothesis, then the linear model is the more parsimonious model to fit the data. Alternatively, when the $TS(G)$ null hypothesis is not rejected, the GM model is the most parsimonious model for the data and the profile is best described by a line with slope 0.

2.4.4. Stratified time-conditional survival probabilities

Consider the initial null hypothesis that the K samples have the same profile and that this same profile shows no change in time-conditional survival probabilities with additional time survived. Under this hypothesis, we expect no differences in adjacent log time-conditional survival estimates and no differences among the strata.

For K-strata, define the p log time-conditional survival probabilities for each of the K populations as

$$\theta_{ki} = \log CS_{ki} = \log\left(\frac{S(b_{ki})}{S(a_{ki})}\right), \qquad i = 1, \ldots, p, \qquad k = 1, \ldots, K.$$

The null and alternative hypotheses are given by

$$H_0 : \log \mathbf{CS} = \mathbf{X}\beta \qquad and \qquad H_1 : \log \mathbf{CS} = \boldsymbol{\theta},$$

where

$$\boldsymbol{\theta} = \left(\theta_{11}, \theta_{12}, \ldots, \theta_{k(p-1)}, \theta_{kp}\right)',$$

is the $Kp \times 1$ vector of all p log time-conditional survival probabilities across K populations. As previously, β is estimated by

$$\hat{\beta} = \left[\mathbf{X}'\hat{\mathbf{\Sigma}}^{-1}\mathbf{X}\right]^{-1} \mathbf{X}'\hat{\mathbf{\Sigma}}^{-1} \log \widehat{\mathbf{CS}},$$

with the $Kp \times 1$ design matrix of ones. Then $\mathbf{X}\hat{\beta}$ represents the single estimated horizontal profile showing no change in time-conditional survival probabilities with additional time survived which is the same across the K samples. The weighted multiple regression test statistic is given by

$$TS = \left(\log \widehat{\mathbf{CS}} - \mathbf{X}\hat{\beta}\right)' \hat{\mathbf{\Sigma}}^{-1} \left(\log \widehat{\mathbf{CS}} - \mathbf{X}\hat{\beta}\right),$$

where $\mathbf{X}$ is the $p \times 1$ vector under the null hypothesis. For large samples, TS is approximately distributed as a central χ^2 with degrees of freedom $Kp - 1$ under the null hypothesis and is distributed as a non-central χ^2 under the alternative hypothesis.

When the null hypothesis is rejected, we evaluate the hypothesis that there is an unspecified relationship between additional time survived and time-conditional survival probabilities, but no difference among the profiles in the K strata. That is, the profiles among the K populations are the same, but that common single profile shows some change in log time-conditional survival probabilities with increasing time after diagnosis. For example, with two strata, this null hypothesis implies that the two log time-conditional survival profiles have the same relationship between additional time survived and time-conditional survival probabilities. The null hypothesis is given by

$$H_0 : \mathbf{X} \log \mathbf{CS} = \begin{pmatrix} \log CS_{11} - \log CS_{21} \\ \vdots \\ \log CS_{1p} - \log CS_{2p} \end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix},$$

a p-length vector, and the alternative hypothesis is given by

$$H_1 : \mathbf{X} \log \mathbf{CS} = \boldsymbol{\theta},$$

where

$$\boldsymbol{\theta} = \left(\theta_{11}, \theta_{12}, \ldots, \theta_{k(p-1)}, \theta_{kp}\right)',$$

is the $Kp \times 1$ vector of all p log time-conditional survival probabilities across K populations. Then, the test statistic is given by

$$TS = \left(\mathbf{X}\hat{\boldsymbol{\theta}}\right)' \left[\mathbf{X}\hat{\mathbf{\Sigma}}\mathbf{X}'\right]^{-1} \left(\mathbf{X}\hat{\boldsymbol{\theta}}\right),$$

where the design matrix has dimensions $p \times Kp$ and $\hat{\boldsymbol{\theta}}$ is a Kp-length vector of the p estimated

log time-conditional survival probabilities for the K strata using the Kaplan-Meier survivor function estimator, $\hat{S}(t)$ (see Equation 2.2). Under the null hypothesis, this test statistic is distributed as χ^2 with $p(K-1)$ degrees of freedom. As an example where $K = 2$ and $p = 3$, the design matrix is given by

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{pmatrix},$$

and the parameter vector is given by

$$\log \mathbf{CS} = (\log CS_{11}, \log CS_{12}, \log CS_{13}, \log CS_{21}, \log CS_{22}, \log CS_{23})'.$$

Using a similar approach, we can consider a constant and fixed strata effect with additional time survived. That is, whatever the relationship is between additional time survived and time-conditional survival probabilities for the one stratum, the others will have a constant increase (or constant decrease) with additional time survived. This results in profiles with the same profile pattern and a fixed difference in magnitude with each additional unit of time survived. Statistically, this implies that the vector of differences of time-conditional survival probabilities will be constant.

2.4.5. Multivariable models for time-conditional survival probabilities

Consider the scenario of two binary variables defined as $X_1 = 0, 1$ and $X_2 = 0, 1$. We are interested in whether independent variable X_1 has a different effect on log time-conditional survival probabilities depending on values of X_2. That is, we are interested in assessing whether there is an interaction effect. Define $k = 4$ strata by $\{X_1, X_2\} \in \{(0,0), (0,1), (1,0), (1,1)\}$ and let $p = 3$, indicating that for each strata there are three log time-conditional survival probability estimators. Under this scenario, the Kp-length vector of log time-conditional survival probabilities is given by

$$\log \mathbf{CS} = (\log CS_{11}, \log CS_{12}, \log CS_{13}, \ldots, \log CS_{41}, \log CS_{42}, \log CS_{43})'.$$

We begin by defining the full saturated model given by

$$E(\log \widehat{\mathbf{CS}}) = \mathbf{X}_f \boldsymbol{\beta}_f,$$

where $\mathbf{X}_f$ is a 12×12 design matrix and $\boldsymbol{\beta}_f$ is a vector of length 12×1. For the full model, the design matrix $\mathbf{X}_f$ is given by

$$\mathbf{X}_f = \begin{pmatrix} \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{I}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{I}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{I}_{3\times3} \end{pmatrix}, \tag{2.13}$$

and $\boldsymbol{\beta}_f$ is given by

$$\boldsymbol{\beta}_f = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{31}, \beta_{32}, \beta_{33})'.$$

Then the expectation for a single log time-conditional survival probability is given by

$$E(\log CS_{kp}) = \beta_{0p} + \beta_{1p} I(X_1 = 1) + \beta_{2p} I(X_2 = 1) + \beta_{3p} I(X_1 = 1) I(X_2 = 1),$$

where $I(\cdot)$ is the indicator function defining the value of $k = 1, 2, 3, 4$. At time point $p = 1, 2, 3$ for this example, the log time-conditional survival probability for each of the four strata as a function of p is given by

$$\begin{array}{rcccccc} E[\log CS_{1p}] & = & \beta_{0p} & & & & \\ E[\log CS_{2p}] & = & \beta_{0p} & +\beta_{1p} & & & \\ E[\log CS_{3p}] & = & \beta_{0p} & & +\beta_{2p} & & \\ E[\log CS_{4p}] & = & \beta_{0p} & +\beta_{1p} & +\beta_{2p} & +\beta_{3p} \end{array}. \tag{2.14}$$

Similarly, for the reduced main effects model, the 12×9 design matrix $\mathbf{X}_r$ is given by

$$\mathbf{X}_r = \begin{pmatrix} \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{I}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} \\ \mathbf{I}_{3\times3} & \mathbf{0}_{3\times3} & \mathbf{I}_{3\times3} \\ \mathbf{I}_{3\times3} & \mathbf{I}_{3\times3} & \mathbf{I}_{3\times3} \end{pmatrix}, \tag{2.15}$$

and the 9×1 vector, $\boldsymbol{\beta}_r$, is given by

$$\boldsymbol{\beta}_r = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{21}, \beta_{22}, \beta_{23})'.$$

For this reduced model, the expectation for a single log time-conditional survival probability is given by

$$E[\log CS_{kp}] = \beta_{0p} + \beta_{1p} I(X_1 = 1) + \beta_{2p} I(X_2 = 1),$$

and, for each group, we have

$$\begin{array}{rcllll} E[\log CS_{1k}] & = & \beta_{0k} & & & \\ E[\log CS_{2k}] & = & \beta_{0k} & +\beta_{1k} & & \\ E[\log CS_{3k}] & = & \beta_{0k} & & +\beta_{2k} & \\ E[\log CS_{4k}] & = & \beta_{0k} & +\beta_{1k} & +\beta_{2k} & \end{array} . \tag{2.16}$$

Now consider the generalization where $X_1, \ldots, X_m$ represent discrete, categorical covariates. Let the multiplicative model include all m covariates and their g interactions. The first consideration is a hypothesis test of whether an interaction term is significant in the saturated model. To test this hypothesis, we can compare the full saturated model to a reduced model without the interaction terms. Then the expectation of log time-conditional survival probability under the full model is given by

$$\begin{aligned} E[\log CS_{ij}] = \beta_0 + \beta_1 I(X_1 = 1) + \beta_2 I(X_2 = 1) + \cdots + \beta_m I(X_m = 1) \\ + \cdots + \beta_g I(X_1 = 1) I(X_2 = 1) \cdots I(X_m = 1), \end{aligned} \tag{Model 1}$$

where $X_1, \ldots, X_m$ are binary variables. Note, g in the full model is determined by the number of the main effects and all of the interaction terms. Define the expectation of log time-conditional survival probability under a reduced model by

$$E[\log CS_{ij}] = \beta_0 + \beta_1 I(X_1 = 1) + \beta_2 I(X_2 = 1) + \cdots + \beta_m I(X_m = 1), \tag{Model 2}$$

which differs from the full saturated model by the exclusion of the interaction terms. Without loss of generality, assume that the $\boldsymbol{\beta}_f$ parameters have been arranged corresponding to the columns of the design matrix, which allows for partitioning. In defining the model in this way, the model parameters from the full model, $\boldsymbol{\beta}_f$, can be partitioned to include the main effects in the reduced model, $\boldsymbol{\beta}_r$, and the parameters from the interaction terms, $\boldsymbol{\beta}_{int}$.

We estimate the parameters, $\boldsymbol{\beta}_f$, in the full saturated model as $\hat{\boldsymbol{\beta}}_f$ with design matrix $\mathbf{X}_f$ using the weighted least squares estimation given by

$$\hat{\boldsymbol{\beta}}_f = \left[\mathbf{X}_f' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_f\right]^{-1} \mathbf{X}_f' \hat{\boldsymbol{\Sigma}}^{-1} \log \widehat{\mathbf{CS}}.$$

Similarly, we estimate the parameters, $\boldsymbol{\beta}_r$, in the reduced model as $\hat{\boldsymbol{\beta}}_r$ with design matrix $\mathbf{X}_r$ by

$$\hat{\boldsymbol{\beta}}_r = \left[\mathbf{X}_r' \hat{\boldsymbol{\Sigma}}^{-1} \mathbf{X}_r\right]^{-1} \mathbf{X}_r' \hat{\boldsymbol{\Sigma}}^{-1} \log \widehat{\mathbf{CS}},$$

for the full and reduced models, respectively. To test the null hypothesis of no interaction, we define the null hypothesis as

$$H_0 : \boldsymbol{\beta}_{int} = \mathbf{0}.$$

We evaluate the null hypothesis that a subset of the parameters from full saturated model with the interaction terms, $\boldsymbol{\beta}_{int}$, are not significant predictors in estimating $\log \mathbf{CS}$ using the test statistic given by

$$TS = \left(\mathbf{X}_f \hat{\boldsymbol{\beta}}_f - \mathbf{X}_r \hat{\boldsymbol{\beta}}_r\right)' \hat{\boldsymbol{\Sigma}}^{-1} \left(\mathbf{X}_f \hat{\boldsymbol{\beta}}_f - \mathbf{X}_r \hat{\boldsymbol{\beta}}_r\right).$$

This test statistic has a χ^2 distribution with $rank(\mathbf{X}_f) - rank(\mathbf{X}_r)$ degrees of freedom under the null hypothesis. If the null hypothesis is rejected, we conclude that the full saturated model cannot be represented by the reduced model with main effects only. In the case where a single interaction or a single parameter such as a main effect is being tested, this becomes a 1 degree of freedom test under the null hypothesis.

2.5. Simulation Studies

Simulations of 10,000 datasets were used to assess the type I error and power. We expect error rates computed from 10,000 datasets will have a 95% confidence interval to be $(0.0457, 0.0543)$. For each dataset, we computed log 5-year time-conditional survival estimates given 0, 1, 2, and 3 years of survival after diagnosis, which are common in the literature (Merrill and Hunter, 2010). Statistics were based on point estimates and the estimated covariance matrix of the log time-conditional

survival probabilities where the profile is given by

$$\log \widehat{\mathbf{CS}} = \left(\log \widehat{CS}_1, \log \widehat{CS}_2, \log \widehat{CS}_3, \log \widehat{CS}_4 \right).$$

Data for the type I error simulations were generated from an exponential distribution, $\exp(\lambda)$. We used a mean parameter $\frac{1}{\lambda} = \frac{1}{7}$ and varied the sample sizes (n = 100, 150, 200, 300, 400, 500, 800, 1000). Three censoring mechanisms were considered: (1) complete data with no censoring, (2) uniform random censoring at 10% in each sample, and (3) uniform random censoring at 35% in each sample. Observed χ^2 test statistics were compared to critical χ^2 values at 3 degrees of freedom for the GM null hypothesis and at 1 degree of freedom for each of the individual contrast hypotheses (α = 0.05).

We determined the parameters of the uniform censoring distribution in order to achieve the specified percentage of censored observations for our simulation study. Let the event time, X, have an exponential distribution with parameter λ, $\exp(\lambda)$, and the censoring time, Y, have a uniform distribution with parameters a and b, $Uniform(a, b)$, where X and Y are independent. The probability of interest is the probability that the censoring time, Y, is less than the event time, X, $P(Y < X)$. Define a random variable transformation where $Z = Y - X$ and $W = Y$. Then the joint distribution of Z and W is given by

$$f_{Z,W}(z, w) = \frac{1}{\lambda} exp\left(-\frac{w - z}{\lambda} \cdot \frac{1}{b - a} \right).$$

For the distribution of Z, integrate the joint distribution over W to get

$$f_Z(z) = \frac{1}{b - a} exp\left(\frac{z}{\lambda}\right) \left(exp\left(-\frac{a}{\lambda}\right) - exp\left(-\frac{b}{\lambda}\right) \right).$$

Using the distribution of Z, we can obtain $P(Y < X)$ or equivalently $P(Z < 0)$ given by

$$P(Z < 0) = \frac{\lambda}{b - a} \left(exp\left(-\frac{a}{\lambda}\right) - exp\left(-\frac{b}{\lambda}\right) \right).$$

For known values of λ, $P(Y < X)$, and $a = 0$, we solve this equation for b. For example, when the time to the event is distributed as $\exp(\lambda = 7)$, the censoring distribution needed to achieve 20% random uniform censoring is distributed as $Uniform(a = 0, b = 34.7558)$.

Data for the power simulations were generated using a Weibull distribution with scale parameter $\lambda = 7$ such that the mean was equal to $1/7$ and a varying shape parameter $\gamma = 0.5, 0.65, 0.8$. Values of λ were chosen to be similar to those observed in the data analysis and were chosen to reflect the SEER data. For this distribution, the probability of surviving beyond 1, 3, 5, and 10 years were 0.8669, 0.6514, 0.4895, and 0.2397, respectively. Values of γ were chosen to be less than one, characterizing a decreasing failure rate over time. These distributions are relevant when mortality is expected to be high after diagnosis and then decrease over time. When $\gamma = 1$, the Weibull distribution simplifies to the exponential distribution with parameter λ. Therefore, the shape parameter of 0.8 was closest to the null distribution of no increasing log time-conditional survival probability with increasing time after diagnosis.

2.5.1. Type I Error

Global Mean and Omnibus Contrast. For the omnibus contrast or global mean test statistic, the results of the type I error simulations are presented in Figure 2.1. In the complete data case, for samples of size 200 and greater, the 95% confidence interval for the estimated type I proportion included the 5% error rate. For uniform random censoring at 10%, the estimated type I error was higher than that for the complete data case. The observed type I error for this test was slightly higher than the 5% error rate for samples of size 100, 150, 200, 300, and 500. For uniform random censoring at 35%, the 95% confidence interval for the observed proportion included 5% only for sample sizes of 800. As the likelihood of uniform random censoring increased, the proportion of test statistics that rejected the true null hypothesis also increased. Overall, with an increased number of censored observations, information was lost and the estimated type I error of the global mean and omnibus contrast tests became less stable and higher than the desired 5% error rate.

Independent Pairwise Contrasts. We defined the first contrast as

$$C_1 = \log \widehat{CS}(5 \mid 0) - \log \widehat{CS}(6 \mid 1),$$

the second contrast as

$$C_2 = \log \widehat{CS}(6 \mid 1) - \log \widehat{CS}(7 \mid 2),$$

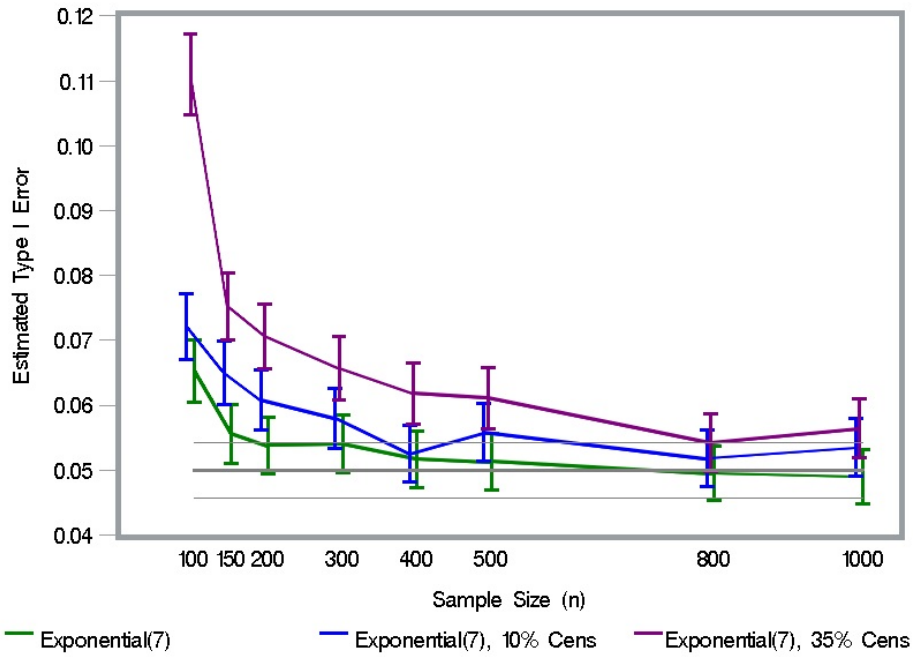

Figure 2.1: Expected type I error of 5% with the 95% confidence interval for 10,000 datasets and estimated type I error for the GM model test statistic (and 95% confidence intervals) for no censoring, 10%, and 35% uniform random censoring.

and the third contrast as

$$C_3 = \log \widehat{CS}(7 \mid 2) - \log \widehat{CS}(8 \mid 3).$$

For the first contrast, C_1, the estimated 95% confidence limits for the observed type I error included 5% when the sample size was 300 and greater. For the second contrast, C_2, the estimated 95% confidence limits for the observed type I error included 5% for all sample sizes considered, except for 300 where it was higher than 0.05 (mean: 0.0571, 95% CI (0.0526, 0.0616)). Lastly, the 95% confidence limits for the observed type I error for the C_3 test statistic consistently include 5% for sample sizes as small as 100.

2.5.2. Power

One-Sample LM and QM. Simulations were used to evaluate the power of the test statistics for the one-sample linear model (LM) and quadratic model (QM). For the LM test, the null hypothesis was that the profile has a linear relationship for log time-conditional survival probability and additional time survived. The alternative hypothesis was that there was no relationship. The power of the LM test was evaluated under an alternative Weibull ($\lambda = 7, \gamma = 0.5$) distribution versus the null

hypothesis of a linear trend. Figure 2.2 shows the power of the LM test for the complete data and for 35% uniform random censoring. For complete data, the LM test achieved 78% power at a sample size of 150. With 35% censoring, we observed only 71% power at that same sample size, although the power increased to 83% when the sample size was 200. For the QM test, the null hypothesis was a quadratic model for the relationship. The alternative hypothesis was that there was no relationship between log time-conditional survival probability and additional time survived. In this situation, much larger sample sizes of $800 - 1000$ were needed to achieve 80% power. With censoring, even larger sample sizes were needed to achieve reasonable power (Figure 2.2). Overall, the statistic evaluating the linear model had consistently higher power than the quadratic model. Under both the LM and QM null hypotheses, uniform random censoring led to a decrease in the observed power of the test.

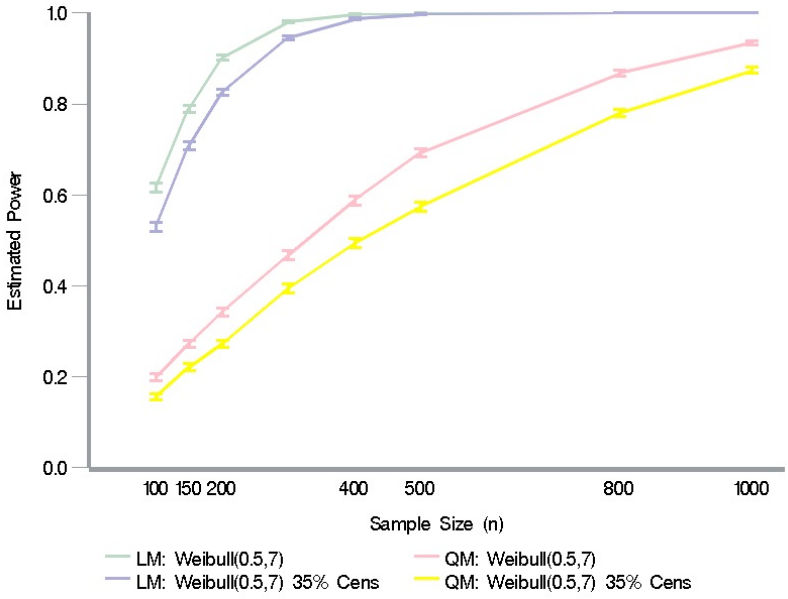

Figure 2.2: Estimated power for the LM and QM test statistics with and without censoring.

Global Mean and Omnibus Contrast. Under the null hypothesis for both the global mean model and the omnibus contrast tests, the more parsimonious fit to the data is a constant profile where there is no change in log time-conditional survival with added time after diagnosis. Figure 2.3 shows the results of the power simulations for the omnibus contrast test statistic and the global mean model test statistic, which both evaluated this null hypothesis of no change. The data generated from the Weibull ($\lambda = 7, \gamma = 0.5$) had a shape parameter that was most different from that of the exponential null distribution ($\gamma = 1$). The power for this test was greater than 80% for sample sizes

100 and greater. The data generated from the Weibull ($\lambda = 7, \gamma = 0.8$) distribution was closest to the null distribution. The power of the test exceeded 78% for sample sizes of 500 and greater. For data generated from the Weibull distribution ($\lambda = 7, \gamma = 0.65$), the power of the global mean model was greater than 90% for sample sizes of 200 and greater.

Under uniform random censoring at 35%, power decreased when compared to no censoring in the complete data case. Although the relationship among the three alternative hypothesis distributions remained the same, there was an exception at sample sizes of 100 for $\gamma = 0.8$ (Figure 2.3). For data where the underlying true distribution was close to the exponential distribution, a larger sample size was required to detect a true difference for the global mean and omnibus contrast tests with 80% power. A larger sample size was required when there was censoring.

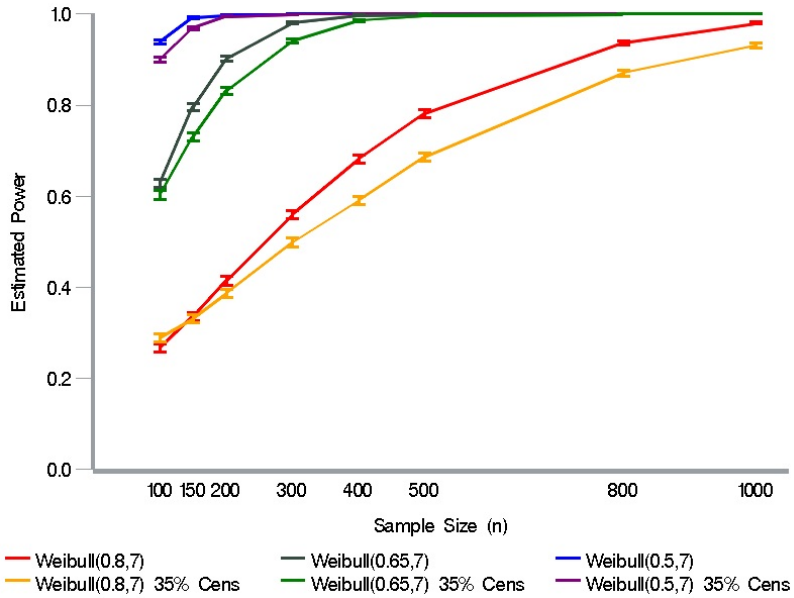

Figure 2.3: Estimated power for the GM model test statistic.

Independent Pairwise Contrasts. For data generated under the Weibull ($\lambda = 7, \gamma = 0.8$) distribution, the test statistic of the C_1 contrast attained 80% power for samples of size 500 and greater. The test for C_2 only attained 45% power by a sample size of 1000 whereas the test for C_3 attained 34% power at that sample size. As the distribution further deviated from the null distribution, under the Weibull ($\lambda = 7, \gamma = 0.65$) 80% power was attained at samples of size 150 and 1000 for tests of the contrasts C_1 and C_2, respectively. The test statistic for the C_3 contrast only attained 51% power by a sample size of 1000. When considering an alternative hypothesis distribution with $\gamma = 0.5$, the

test of contrast C_1 attained 96% power at a sample of size 100 and C_2 attained 78% power for a sample of size 500. The test of the last contrast, C_3, reached 71% power at a sample of size 1000. Similar to the global mean and omnibus contrast tests, we found a larger sample size was required to achieve power above 80% as the alternative hypothesis distribution approached the exponential null distribution.

2.6. Application: Staging Procedure and Time-Conditional Survival Probability for Stage II Melanoma Patients

We illustrate our time-conditional survival methodology using population-based data on patients with cutaneous melanoma from the Surveillance, Epidemiology, and End Results (SEER) program (SEER, 2008). Our study included 5370 patients diagnosed in 2004–2008 with a single primary melanoma and no palpable regional lymph nodes. These patients had Stage II disease defined by the SEER Derived AJCC Stage Group variable (derived from the Collaborative Stage detailed site-specific codes, using the Collaborative Stage algorithm) with or without lesional ulceration (Balch et al., 2001, 2009). To obtain this study sample, we applied inclusion and exclusion criteria. Inclusion criteria included: (1) primary site defined by ICD-O-2 coding under other malignant neoplasms of skin (C44.0, C44.1, C44.2, C44.3, C44.4, C44.5, C44.6, C44.7, C44.8, C44.9), (2) histologic type ICD-O-3 code describing the microscopic composition of the cells and/or tissue for a specific primary ranging 8720 to 8799, (3) first malignant primary tumor using the behavior ICD-O-3 code indicating a malignant primary site (invasive) tumor, (4) diagnostic confirmation microscopically confirmed (positive histology or positive microscopic confirmation with the method not specified), (5) survival time greater than 0 months, (6) exactly one primary lesion, (7) disease diagnosis in years 2004-2008 inclusive, and (8) Derived AJCC Stage Group variable indicating Stage II disease (including NOS, IIA, IIB, and IIC). Exclusion criteria included: (1) unknown ulceration status and (2) unknown status, scope or pathological findings at regional lymph node surgery.

In this SEER-based study, patients were analyzed both as a single cohort and as multiple cohorts classified by two variables: ulceration (yes or no) and staging (clinical or pathological). As in Gimotty et al., 2005, we assume that, if no nodes were surgically sampled, patients had no evidence of nodal involvement by palpation (they were thus *clinically staged* as II). If they had a surgical procedure(s) to examine their regional nodes they were *pathologically staged* as II when no nodes

had histological evidence of metastases. The kind of nodal evaluation and its results are important because of its relation to staging. In the AJCC staging system (Balch et al., 2001), the absence or presence of metastases in regional nodes as determined by palpation or by a surgical procedure is a key component of the TNM classification. In the SEER data, the pathological N classification is determined by the number of regional lymph nodes removed and the number of positive lymph nodes found. Similarly, ulceration of the primary lesion is one of the prognostic factors used in the AJCC staging system. Ulceration is defined by SEER as the "absence of an intact epidermis overlying the primary melanoma based upon histopathological examination" (SEER, 2008). Patients were classified into two groups defined by the presence or absence of ulceration.

The patients included in this study were stage II subjects who had their melanomas apparently confined to the primary site with no evidence of palpable regional lymph nodes. In our data, differences in patients may arise due to a mixture of clinical and pathological staging and a mixture of ulceration status. Clinically staged patients have only the absence of papably enlarged nodes to indicate the absence of metastases; pathologically staged patients had this and histological evidence of uninvolved nodes. We hypothesize that this heterogeneity will result in patients with clinical stage II disease (who may have small amounts of metastatic disease in their regional nodes) having an inferior survival experience to those with pathological stage II disease.

For patients presenting with clinical stage II melanoma, guidelines recommend that a surgical procedure called sentinel lymph node biopsy (SLNB) be offered to more accurately identify patients with clinically not apparent (not palpable) metastasis to regional lymph nodes (Morton et al., 2006). This is a powerful staging and prognostic procedure because it allows identification of metastases that are not clinically observable. As SLNB is not always performed, we were able to examine patients who had either clinical or pathological staging of disease. Three-year time-conditional survival probabilities were defined for time from diagnosis to melanoma-specific death estimated from the Kaplan-Meier survival curves at 6, 12, and 18 months from diagnosis. We sought to investigate patterns of 3-year time-conditional survival in patients who had clinical (no nodal procedure) or pathological (some nodal procedure) evaluation of their regional nodes, controlling for ulceration (Gamerman et al., 2012). Future survival, beyond the time already survived, was captured by estimating time-conditional survival probabilities.

For testing hypotheses related to a single profile, the first analysis approach evaluated changes

in time-conditional survival as the time survived after diagnosis increased. The second analysis approach investigated whether the relationship between time-conditional survival and increasing time survived was different for patients with different types of tumors (ulcerated and not ulcerated) and different types of nodal evaluation (clinical or pathological).

2.6.1. Estimation of a single time-conditional survival profile

We estimated time-conditional survival probabilities and their 95% confidence intervals (Equations 2.5 and 2.6) for all of the patients in our cohort with and without ulcerated lesions who had clinical or pathological nodal staging. We found that 3-year time-conditional survival probabilities decreased in the first 6 months after diagnosis, plateaued between months 6 and 12 after diagnosis, and then increased 18 months after diagnosis. In Figure 2.4, the 3-year log time-conditional survival probability estimates and their 95% confidence intervals were -0.132 (-0.146, -0.118) at diagnosis, -0.143 (-0.159, -0.128) at 6 months after diagnosis, -0.144 (-0.162, -0.126) at 12 months after diagnosis, and -0.128 (-0.147, -0.109) at 18 months after diagnosis. An analysis based only on confidence intervals would be unable to show definitively any difference in the point estimates because their confidence intervals overlap (Figure 2.4). This would imply that there was no difference in the estimates and that additional time after diagnosis was not associated with changes in estimates of 3-year log time-conditional survival probabilities. Appropriate simultaneous inference on multiple estimates using data from the entire time-conditional survival profile would require their joint distribution and would account for the correlation in point estimates.

2.6.2. Omnibus test of contrasts for a single profile

Using the omnibus contrast test, the null hypothesis was rejected, suggesting there were differences between adjacent 3-year log time-conditional survival probabilities with additional time survived ($p = 0.0004$). We investigated this relationship using independent pairwise contrasts. After using the Bonferroni adjustment for multiple comparisons, two of the three tests were statistically significant (Table 2.1). There was statistically significant change in 3-year log time-conditional survival probability between evaluation at diagnosis and 6 months ($\log CS_1 - \log CS_2$, $p = 0.0075$), as well as a change between evaluation at 12 months and 18 months ($\log CS_3 - \log CS_4$, $p = 0.0072$). The p-values were the Bonferroni-adjusted p-values corresponding to the unadjusted p-values in Table 2.1. Therefore, the drop from -0.132 to -0.143 between diagnosis and 6 months demonstrated

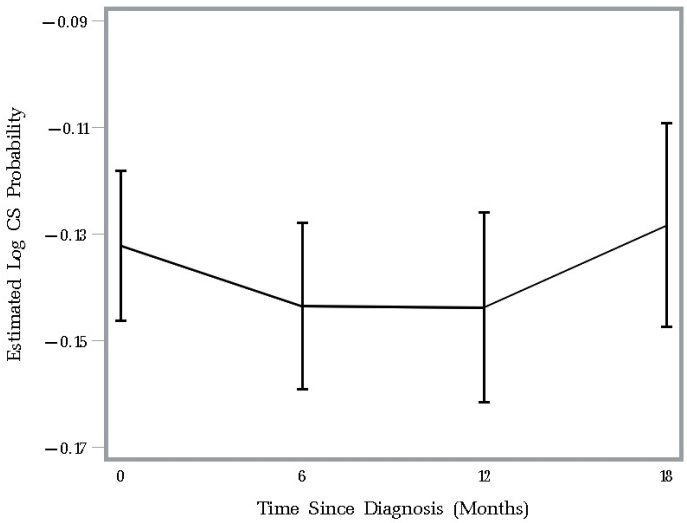

Figure 2.4: Nonparametric Kaplan-Meier estimated 3-year log time-conditional survival probabilities given 0, 6, 12, and 18 months after diagnosis for Stage II patients.

a statistically significant decline in the 3-year log time-conditional survival probability. Additionally, the observed change in the estimates from 12 to 18 months after diagnosis indicated a statistically significant increase in the 3-year log time-conditional survival probability at -0.144 and -0.128, respectively (Figure 2.4).

In our cohort, the observation of an initial drop in log time-conditional survival probability estimates, followed by an increase in these estimates was not surprising. The pattern in traditional Kaplan-Meier survival of a decreasing failure rate over time, such that there was higher mortality after diagnosis and then decreasing mortality over time, has been regularly observed for melanoma patients (e.g., Balch et al., 2010; Xing et al., 2010). Clinical and pathological Stage II patients included those with melanomas with classification N0, M0, T2b, T3a, T3b, T4a, and T4b where N0 represented no apparent regional lymph node metastasis and M0 represented no apparent distant (beyond the regional nodes) metastasis. The primary tumor (T) classification represented: (1) melanoma 1.01 - 2.0 mm in thickness with ulceration (T2b), (2) melanoma 2.01 - 4.0 mm in thickness with or without ulceration (T3), and (3) melanoma greater than 4.0 mm in thickness with or without ulceration (T4). This variability in primary tumor thickness and ulceration of the primary tumor were related to differences in observed survival (Balch et al., 2009).

Looking at 3-year estimates of time-conditional survival at diagnosis and 6, 12, and 18 months after diagnosis in this study, we found that 3-year time-conditional survival at 6 months after diagnosis

declined. After that, the outlook for future prognosis of 3 additional years begins to improve with each additional 6 months.

2.6.3. Fitting a parsimonious regression model to the profile

To model the 3-year time-conditional survival probabilities, we fit the QM, LM, and GM models. The parameter estimates from these three models are presented in Table 2.2. Under the alternative H_1, the estimated log time-conditional survival probabilities varied from -0.144 to -0.128. Also presented is the estimated covariance matrix of the estimated log time-conditional survival probabilities. The variances are shown down the diagonal, with the covariances in the upper triangle, and the correlations in the lower triangle. Estimates of log-time conditional survival probabilities for this profile were positively correlated ($0.39 - 0.95$).

We found that the quadratic model was the most parsimonious model for these data ($TS = 0.14, p = 0.7107$). We could not simplify the model to the linear model ($TS = 18.01, p = 0.0001$) and global mean model ($TS = 18.36, p = 0.0004$ for both). Comparing confidence intervals without accounting for the correlation of the parameters would have concluded that increasing time after diagnosis did not influence time-conditional survival.

2.6.4. Estimation of stratified time-conditional survival profiles

We were interested in whether patients with and without nodal procedure and with and without ulcerated primary tumors would have different time-conditional survival profiles. The nonparametric Kaplan-Meier log time-conditional survival probability estimates plotted against time since diagnosis for the two independent groups defined by nodal procedure (clinical or pathological) are shown in Figure 2.5 and ulceration status (not ulcerated or ulcerated) are shown in Figure 2.6. Figure 2.7 shows the nonparametric Kaplan-Meier log time-conditional survival probability estimates plotted against time since diagnosis for the four independent groups defined by the two binary covariates, nodal procedure (clinical or pathological) and ulceration status (present or absent). The objective was to evaluate the relationship between 3-year time-conditional survival probability and additional time survived after diagnosis for these groups.

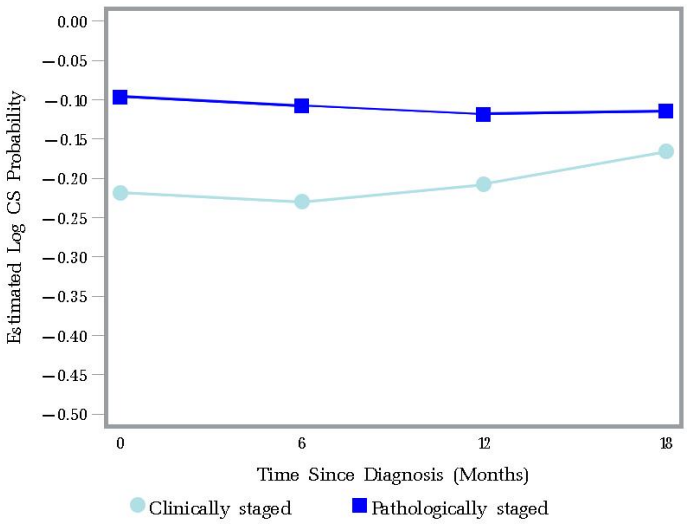

Figure 2.5: Nonparametric Kaplan-Meier estimated 3-year log time-conditional survival probabilities given 0, 6, 12, and 18 months after diagnosis for patients by procedure: (1) Pathologically staged (some nodal procedure) and (2) Clinically staged (no nodal procedure).

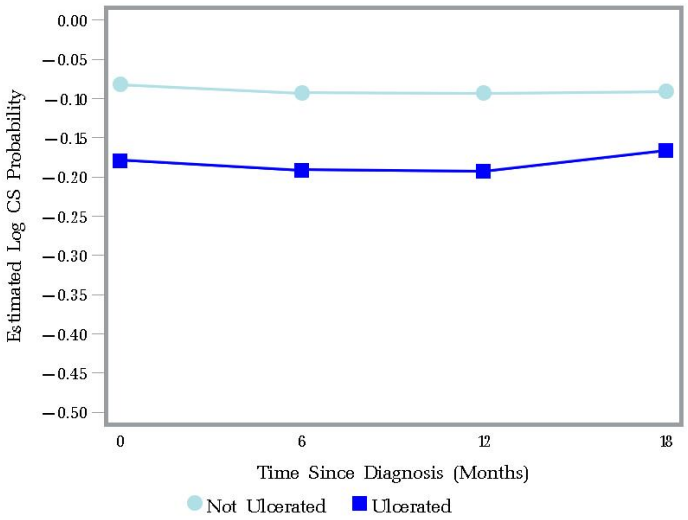

Figure 2.6: Nonparametric Kaplan-Meier estimated 3-year log time-conditional survival probabilities given 0, 6, 12, and 18 months after diagnosis for patients by ulceration status: (1) Not ulcerated and (2) Ulcerated.

2.6.5. Stratified time-conditional survival probabilities comparing two groups

As examples of the use of the two-sample methodology, we considered nodal procedure (Figure 2.5) and ulceration status (Figure 2.6) independently. Under the null hypothesis of no strata effect and no effect of additional time after diagnosis, we would observe the two profiles overlapping and no change in log time-conditional survival probabilities with additional months survived

after diagnosis. Let $\mathbf{X}$ is a 16×1 vector of ones. The null and alternative hypotheses are given by

$$H_0 : \log \mathbf{CS} = \mathbf{X}\beta \qquad \text{and} \qquad H_1 : \log \mathbf{CS} = \boldsymbol{\theta}.$$

The parameter β is estimated by $\hat{\beta}$ where

$$\hat{\beta} = \left[\mathbf{X}'\hat{\mathbf{\Sigma}}^{-1}\mathbf{X}\right]^{-1} \mathbf{X}'\hat{\mathbf{\Sigma}}^{-1} \log \widehat{\mathbf{CS}}.$$

Under the null hypothesis, the parameter estimate, $\hat{\beta}$ was -0.110. We concluded that there was evidence to reject the null hypothesis of no nodal procedure strata effect and no change in log time-conditional survival probabilities ($p < 0.0001$). The absolute difference between the two strata was 0.122, 0.123, 0.090, and 0.052 at diagnosis, and at 6, 12, and 18 months after diagnosis, respectively.

Similarly, we considered ulceration status of the tumor at diagnosis. Under the null hypothesis, the parameter estimate, $\hat{\beta}$ was -0.112. The absolute observed difference between the two strata shown in Figure 2.6 was 0.096, 0.098, 0.099, and 0.074 at diagnosis, and at 6, 12, and 18 months after diagnosis, respectively. For ulceration status, we rejected the null hypothesis of no effect of either time after diagnosis or ulceration status and concluded that there is at least one non-zero difference in probabilities ($p < 0.0001$).

To better understand the observed data, we considered the null hypothesis of no strata effect but allowed for an effect due to additional time survived after diagnosis. The null and alternative hypotheses are given by

$$H_0 : \begin{pmatrix} \log CS_{11} - \log CS_{21} \\ \log CS_{12} - \log CS_{22} \\ \log CS_{13} - \log CS_{23} \\ \log CS_{14} - \log CS_{24} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

and

$$H_1 : \begin{pmatrix} \log CS_{11} - \log CS_{21} \\ \log CS_{12} - \log CS_{22} \\ \log CS_{13} - \log CS_{23} \\ \log CS_{14} - \log CS_{24} \end{pmatrix} = \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \\ \Delta_4 \end{pmatrix},$$

where $\log CS_{kp}$ represents the log time-conditional survival estimator in the k-th strata at the p-th time already survived. Under this null hypothesis, we would observe the strata profiles overlapping, however, the log time-conditional survival profile could increase or decrease with additional time survived. The alternative hypothesis is that there exists some non-zero difference in log time-conditional survival probabilities. For these data, we had evidence to reject the null and concluded that there was at least one significant non-zero difference log time-conditional survival probabilities ($p < 0.0001$ for nodal procedure). Similarly, we considered ulceration status of the tumor at diagnosis. For ulceration status, we rejected the null hypothesis of no effect of the ulceration strata. We concluded that there was a significant non-zero difference in at least one time-conditional survival probability ($p < 0.0001$).

Allowing for a relationship between time survived and 3-year log time-conditional survival probabilities, we considered the hypothesis of a fixed strata effect. The null and alternative hypotheses are given by

$$H_0 : \begin{pmatrix} \log CS_{11} - \log CS_{21} \\ \log CS_{12} - \log CS_{22} \\ \log CS_{13} - \log CS_{23} \\ \log CS_{14} - \log CS_{24} \end{pmatrix} = \begin{pmatrix} \Delta \\ \Delta \\ \Delta \\ \Delta \end{pmatrix},$$

and

$$H_1 : \begin{pmatrix} \log CS_{11} - \log CS_{21} \\ \log CS_{12} - \log CS_{22} \\ \log CS_{13} - \log CS_{23} \\ \log CS_{14} - \log CS_{24} \end{pmatrix} = \begin{pmatrix} \Delta_1 \\ \Delta_2 \\ \Delta_3 \\ \Delta_4 \end{pmatrix} = \boldsymbol{\Delta}.$$

This null hypothesis states that, when comparing the strata of patients that are clinically or pathologically staged, the difference in 3-year log time-conditional survival probability is constant, independent of how much time after diagnosis has passed (represented by Δ in the null hypothesis above).

Under the null hypothesis of a fixed group effect, the estimated fixed difference in estimated log time-conditional survival probabilities between the two strata was $\hat{\Delta} = -0.095$. We rejected the null hypothesis and found that at least one difference in log time-conditional survival probabilities between the two groups was significantly different from the fixed difference ($p < 0.0001$). We concluded that, while 3-year estimates changed with additional time survived, the significant strata effect found in the previous hypothesis test was not a fixed strata effect for the two groups of patients by nodal procedure. Similarly, to evaluate if there was a fixed difference in 3-year log time-conditional survival probabilities for patients with ulcerated and non-ulcerated lesions, we considered the null hypothesis that there was a fixed difference in their estimates with increased time survived after diagnosis. We failed to reject the null hypothesis that there was a fixed effect of ulceration status estimated to be $\hat{\Delta} = 0.091$ ($p = 0.1128$). We concluded that patients with ulcerated lesions had poorer 3-year time-conditional survival prognoses as compared to patients with non-ulcerated lesions. This difference was constant with additional time after diagnosis.

2.6.6. Omnibus test of contrasts by stratum

Next, we compared the estimated time-conditional survival probabilities for the four independent profiles defined as (1) pathologically staged and not ulcerated ($n = 1702$), (2) pathologically staged and ulcerated ($n = 1915$), (3) clinically staged and not ulcerated ($n = 801$), and (4) clinically staged and ulcerated ($n = 952$) (Figure 2.7). With increasing time survived up to 18 months, there was no significant change in the 3-year log time-conditional survival estimates for the cohort of pathologically staged patients with non-ulcerated lesions. For pathologically staged patients with ulcerated lesions, the 3-year time-conditional survival estimates changed significantly with increased time after diagnosis under the omnibus contrast test ($p = 0.0089$). None of the independent pairwise contrasts were significantly different from zero under the conservative Bonferroni adjustment ($p_{C_1} = 0.0685$, $p_{C_2} = 0.1813$, $p_{C_3} = 0.0.2651$).

For the cohort of clinically staged patients with non-ulcerated lesions, 3-year log time-conditional survival increased from -0.162 at 12 months after diagnosis to -0.136 at 18 months after diagnosis ($p = 0.0010$). On the time-conditional survival scale, the estimates increased from 85% to 87% at 12 and 18 months after diagnosis, respectively. Lastly, for clinically staged patients with ulcerated lesions, the 3-year log time-conditional survival estimates for those who had survived 6, 12, or 18 months were -0.296, -0.251, -0.195, respectively. On the time-conditional survival scale, the

estimates were 74%, 78%, and 82%, respectively. After the initial 6 months after diagnosis, there was a significant change in 3-year log time-conditional survival having survived 6 versus 12 months ($p = 0.0044$), as well as having survived 12 versus 18 months ($p < 0.001$).

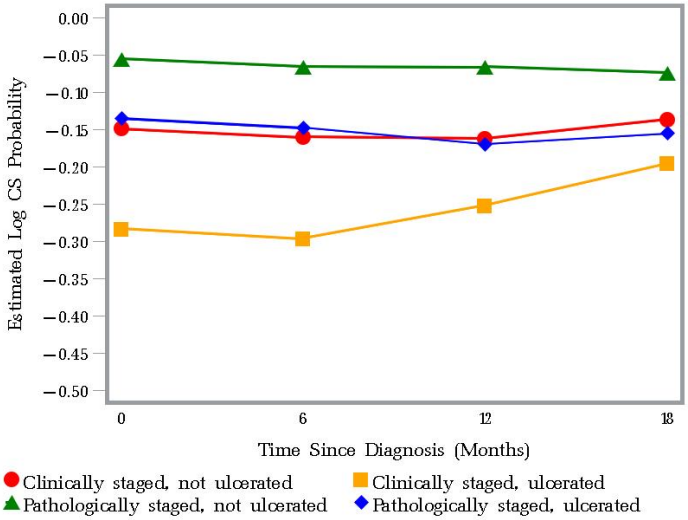

Figure 2.7: Nonparametric Kaplan-Meier estimated 3-year log time-conditional survival probabilities given 0, 6, 12, and 18 months after diagnosis for patients in one of four groups: (1) Pathologically staged (some procedure) and not ulcerated, (2) Pathologically staged (some procedure) and ulcerated, (3) Clinically staged (no nodal procedure) and not ulcerated, and (4) Clinically staged (no nodal procedure) and ulcerated.

Based on the estimated curves, for pathologically staged patients, we concluded that 3-year log time-conditional survival did not change significantly with increasing time after diagnosis from 0 to 18 months. Clinically staged patients with non-ulcerated lesions had an increase in log time-conditional survival from 12 to 18 months after diagnosis, while those with ulcerated lesions had an increase in estimates as early as 6 months after diagnosis. Patients with pathologically staged disease (some procedure) had better log time-conditional survival probability, indicative of better survival, as compared to patients with clinically staged disease (no nodal procedure). Given these individual results, we proceeded to build a multivariable model to assess the relationship between 3-year time-conditional survival probability and additional time survived after diagnosis as a function of the nodal procedure and ulceration status.

2.6.7. Multivariable analysis

In the individual profiles, we observed differential profiles based on the covariate patterns. To determine if the impact of nodal procedure on estimates of log time-conditional survival proba-

bility depended on ulceration status (or, similarly, if the impact of ulceration status on estimates of log time-conditional survival probability depended on nodal procedure), we built a model for log time-conditional survival adjusting for these covariates and their interaction. Four independent groups were defined by two binary covariates, X_1 for nodal procedure (pathologically-staged versus clinically-staged) and X_2 for ulceration status (ulcerated versus not ulcerated). Let the $k = 4$ strata be defined by $\{X_1, X_2\} \in \{(0,0),(0,1),(1,0),(1,1)\}$ such that they represent: (1) clinically-staged patients with no ulceration of the lesion, (2) clinically-staged patients with an ulcerated lesion, (3) pathologically-staged patients with no ulceration of the lesion, and (4) pathologically-staged patients with an ulcerated lesion. Let $p = 4$ indicate that for each stratum of patients there are 4 log time-conditional survival probability estimates defined by 3-year log time-conditional survival probability given survival at diagnosis, 6, 12, and 18 months after diagnosis.

Define the full saturated model given by

$$E[\log \widehat{\mathbf{CS}}] = \mathbf{X}_f \boldsymbol{\beta}_f,$$

where $\mathbf{X}_f$ is a 16×16 design matrix and $\boldsymbol{\beta}_f$ is a vector of length 16×1. For the full model, the design matrix $\mathbf{X}_f$ is given by

$$\mathbf{X}_f = \begin{pmatrix} \mathbf{I}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} \\ \mathbf{I}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{I}_{4\times4} & \mathbf{0}_{4\times4} \\ \mathbf{I}_{4\times4} & \mathbf{I}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} \\ \mathbf{I}_{4\times4} & \mathbf{I}_{4\times4} & \mathbf{I}_{4\times4} & \mathbf{I}_{4\times4} \end{pmatrix},$$

and $\boldsymbol{\beta}_f$ is given by

$$\boldsymbol{\beta}_f = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24}, \beta_{31}, \beta_{32}, \beta_{33}, \beta_{34})'.$$

Then the expectation for a single log time-conditional survival probability is given by

$$\log CS_{kp} = \beta_{0p} + \beta_{1p} I(X_1 = 1) + \beta_{2p} I(X_2 = 1) + \beta_{3p} I(X_1 = 1) I(X_2 = 1), \qquad \text{(Model 1*)}$$

where $I(\cdot)$ is the indicator function defining the value of $k = 1, 2, 3, 4$, and at time point $p = 1, 2, 3, 4$.

For this data, the log time-conditional survival probability for each of the 4 strata as a function of p is given by

$$\begin{array}{rcccccc} \log CS_{1p} & = & \beta_{0p} & & & & \\ \log CS_{2p} & = & \beta_{0p} & & & +\beta_{2p} & \\ \log CS_{3p} & = & \beta_{0p} & +\beta_{1p} & & & \\ \log CS_{4p} & = & \beta_{0p} & +\beta_{1p} & +\beta_{2p} & +\beta_{3p} & \end{array}.$$

We were interested in whether the interaction effect, $\boldsymbol{\beta}_3 = (\beta_{31}, \beta_{32}, \beta_{33}, \beta_{34})'$, differed significantly from a four dimensional zero vector. To test this hypothesis, we defined the reduced model

$$E[\log \widehat{\mathbf{CS}}] = \mathbf{X}_r \boldsymbol{\beta}_r,$$

where $\mathbf{X}_r$ is a 12×12 design matrix and $\boldsymbol{\beta}_r$ is a vector of length 12×1. For this reduced model, the design matrix $\mathbf{X}_r$ is given by

$$\mathbf{X}_r = \begin{pmatrix} \mathbf{I}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{0}_{4\times4} \\ \mathbf{I}_{4\times4} & \mathbf{0}_{4\times4} & \mathbf{I}_{4\times4} \\ \mathbf{I}_{4\times4} & \mathbf{I}_{4\times4} & \mathbf{0}_{4\times4} \\ \mathbf{I}_{4\times4} & \mathbf{I}_{4\times4} & \mathbf{I}_{4\times4} \end{pmatrix},$$

and $\boldsymbol{\beta}_r$ is given by

$$\boldsymbol{\beta}_r = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24})'.$$

Then the expectation for a single log time-conditional survival probability is given by

$$\log CS_{kp} = \beta_{0p} + \beta_{1p} I(X_1 = 1) + \beta_{2p} I(X_2 = 1). \qquad \text{(Model 2*)}$$

At time point $p = 1, 2, 3, 4$, the log time-conditional survival probability for each of the $k = 1, 2, 3, 4$ strata as a function of p is given by

$$\begin{array}{rcccccc} \log CS_{1p} & = & \beta_{0p} & & & & \\ \log CS_{2p} & = & \beta_{0p} & & & +\beta_{2p} & \\ \log CS_{3p} & = & \beta_{0p} & +\beta_{1p} & & & \\ \log CS_{4p} & = & \beta_{0p} & +\beta_{1p} & +\beta_{2p} & & \end{array}.$$

We evaluated this null hypothesis by comparing the full model (Model 1*) with the reduced model (Model 2*). Table 2.3 presents parameter estimates of $\boldsymbol{\beta}_f$ and $\boldsymbol{\beta}_r$ and Table 2.4 represents the estimates of log time-conditional survival probabilities under each model. We found that $\boldsymbol{\beta}_3$ was not significant (p=0.0621). We did not reject the null hypothesis of no interaction at the $\alpha = 0.05$ level.

Next, we considered removing one of the main effects from Model 2*. To test whether ulceration status is significant in the model, we defined the first main effect model as

$$E[\log \widehat{\mathbf{CS}}] = \mathbf{X}_A \boldsymbol{\beta}_A,$$

where $\mathbf{X}_A$ is a 8×8 design matrix and $\boldsymbol{\beta}_A$ is a vector of length 8×1. For this main effect model, the design matrix $\mathbf{X}_A$ is given by

$$\mathbf{X}_A = \begin{pmatrix} \mathbf{I}_{4\times4} & \mathbf{0}_{4\times4} \\ \mathbf{I}_{4\times4} & \mathbf{0}_{4\times4} \\ \mathbf{I}_{4\times4} & \mathbf{I}_{4\times4} \\ \mathbf{I}_{4\times4} & \mathbf{I}_{4\times4} \end{pmatrix}, \tag{2.17}$$

and $\boldsymbol{\beta}_A$ is given by

$$\boldsymbol{\beta}_A = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{11}, \beta_{12}, \beta_{13}, \beta_{14})'.$$

Then the expectation for a single log time-conditional survival probability is given by

$$\log CS_{kp} = \beta_{0p} + \beta_{1p} I(X_1 = 1). \tag{Model 3*}$$

At time point $p = 1, 2, 3, 4$, the log time-conditional survival probability for each of the $k = 1, 2, 3, 4$ strata as a function of p is given by

$$\begin{array}{rcll} \log CS_{1p} & = & \beta_{0p} & \\ \log CS_{2p} & = & \beta_{0p} & \\ \log CS_{3p} & = & \beta_{0p} & +\beta_{1p} \\ \log CS_{4p} & = & \beta_{0p} & +\beta_{1p} \end{array} . \tag{2.18}$$

We evaluated the null hypothesis $\boldsymbol{\beta}_2 = \mathbf{0}$ by comparing the reduced model (Model 2*) with the main

effects Model 3*. Table 2.3 presents parameter estimates of $\boldsymbol{\beta}_f$ and $\boldsymbol{\beta}_r$, and Table 2.4 represents the estimates of log time-conditional survival probabilities under each model. We found that $\boldsymbol{\beta}_2$ was significant ($p < 0.0001$). We rejected the null hypothesis of no effect of ulceration status at the $\alpha = 0.05$ level.

Similarly, we considered removing the second of the main effects from Model 2*. To test whether nodal procedure status was significant in the model, we defined the second main effect model as

$$E[\log \widehat{\mathbf{CS}}] = \mathbf{X}_B \boldsymbol{\beta}_B,$$

where $\mathbf{X}_B$ is a 8×8 design matrix and $\boldsymbol{\beta}_B$ is a vector of length 8×1. For this main effect model, the design matrix $\mathbf{X}_B$ is given by

$$\mathbf{X}_B = \begin{pmatrix} \mathbf{I}_{4\times4} & \mathbf{0}_{4\times4} \\ \mathbf{I}_{4\times4} & \mathbf{I}_{4\times4} \\ \mathbf{I}_{4\times4} & \mathbf{0}_{4\times4} \\ \mathbf{I}_{4\times4} & \mathbf{I}_{4\times4} \end{pmatrix}, \tag{2.19}$$

and $\boldsymbol{\beta}_B$ is given by

$$\boldsymbol{\beta}_B = (\beta_{01}, \beta_{02}, \beta_{03}, \beta_{04}, \beta_{21}, \beta_{22}, \beta_{23}, \beta_{24})'.$$

Then the expectation for a single log time-conditional survival probability is given by

$$\log CS_{kp} = \beta_{0p} + \beta_{2p} I(X_1 = 1). \qquad \text{(Model 4*)}$$

At time point $p = 1, 2, 3, 4$, the log time-conditional survival probability for each of the $k = 1, 2, 3, 4$ strata as a function of p is given by

$$\begin{array}{rcll} \log CS_{1p} & = & \beta_{0p} & \\ \log CS_{2p} & = & \beta_{0p} & +\beta_{2p} \\ \log CS_{3p} & = & \beta_{0p} & \\ \log CS_{4p} & = & \beta_{0p} & +\beta_{2p} \end{array} . \tag{2.20}$$

We evaluated the null hypothesis $\boldsymbol{\beta}_1 = \mathbf{0}$ by comparing the reduced model (Model 2*) with the main

effects Model 4*. Table 2.3 presents parameter estimates of $\boldsymbol{\beta}_f$ and $\boldsymbol{\beta}_r$, and Table 2.4 represents the estimates of log time-conditional survival probabilities under each model. We found that $\boldsymbol{\beta}_1$ was significant ($p < 0.0001$). We rejected the null hypothesis of no effect of nodal procedure at the $\alpha = 0.05$ level.

We found that the most parsimonious model for these data included both main effects, but not the interaction parameter. We concluded that both nodal procedure and ulceration status had significant impact on estimates of log time-conditional survival probability. We built a model for log time-conditional survival probability adjusting for these covariates (Model 2*), where we defined four independent groups based on the two binary covariates. Table 2.3 shows the estimates resulting from this model.

Three-year log time-conditional survival probabilities were estimated at diagnosis, 6, 12, and 18 months after diagnosis. We found that, for clinically staged patients with no ulceration of the lesion, $(k = 1)$, 3-year time-conditional survival probabilities started off lower and increased with increasing time (probabilities ranging from 0.839 to 0.884). A similar pattern was observed for clinically staged patients with an ulcerated lesion, $(k = 2)$, with probabilities ranging from 0.767 to 0.825. For patients with pathologically staged melanoma with no ulceration of the lesion, $(k = 3)$, time-conditional survival probabilities were higher at diagnosis and decreased with increasing time after diagnosis (probabilities decreased from 0.948 to 0.929). Lastly, for pathologically staged patients with an ulcerated lesion, $(k = 4)$, time-conditional survival probability fluctuated about 0.86 (probabilities ranging from 0.853 to 0.869). Overall, the range in time-conditional survival probabilities was more than 2 times narrower for pathologically staged patients (0.019 for k=3 and 0.016 for k=4) as compared to clinically staged patients (0.045 for k=1 and 0.058 for k=2).

2.7. Discussion

In this chapter, we proposed methods for investigating time-conditional survival probabilities as a function of increasing time survived after diagnosis. Estimates of time-conditional survival probabilities with 95% confidence intervals are commonly reported in medical literature. While standalone estimated probabilities have a straightforward interpretation, the developed methodology is able to address clinically relevant questions of interest such as “Does the expected probability that a patient will survive an additional 5 years *significantly* increase with increasing time post-diagnosis”?

To evaluate these relationships, we derived the asymptotic joint distribution of the estimated log time-conditional survival probabilities and developed hypothesis tests to address questions of interest.

To explore whether there are differences in adjacent pairwise log time-conditional survival probabilities, we developed an omnibus test and subsequent independent contrasts to identify pairwise differences. This is an important issue when evaluating a single time-conditional survival probability profile for both researcher and patient because a constant profile indicates that increasing time survived after diagnosis does not imply an increased (or decreased) likelihood of surviving an additional number of years. On the other hand, determining that a profile is characterized by a linear or quadratic relationship provides additional prognostic information to the relationship between time-conditional survival probabilities and additional time survived. We proposed three nested hypothesis tests using the weighted least squares regression approach for a profile-based analysis to identify linear and quadratic relationships between log time-conditional survival probabilities and time survived after diagnosis.

The profile-based methods were extended by fitting models for log time-conditional survival probability profiles incorporating covariate information. Our regression modeling strategy can be used to addresses questions about factors that affect the profiles, by comparing a particular (full) model with a more parsimonious nested (reduced) model. Our regression framework can be used to compare time-conditional survival probabilities from independent populations. Lastly, when profiles are determined from several categorical variables to create independent strata using covariate patterns, we consider a multivariable multiplicative regression model framework to model the relationships between covariates and time-conditional survival probabilities, including the assessment of interactions between covariates and main effects. More detailed investigations into the influence of continuous, as opposed to categorical, prognostic and treatment factors may add utility for monitoring and understanding changes in time-conditional survival probabilities.

In simulations, we demonstrated that the test statistics for the global mean model had good statistical properties for samples of size 200 or greater. As the percentage of uniform censoring increases from 0% to 35%, the sample size necessary for adequate type I error increases. The sample size necessary to achieve adequate power also increased with an increasing percentage of uniform random censoring. The LM test achieved adequate power for a sample size of 200, while the QM

test needed much larger sample sizes to achieve power of 80%. Thus, one limitation of the current hypothesis testing framework is the requirement of large sample sizes to ensure the asymptotic properties of the test statistics.

These methods were applied to real-world survival data from SEER for patients with melanoma. We found that the quadratic model was the most parsimonious model for these data since we could not simplify it to a linear or global mean models ($p < 0.0001$ for both). Note that a limitation of this application was the short interval of time from diagnosis (6, 12, and 18 months) when considering 3-year log time-conditional survival probabilities. Comparing confidence intervals without accounting for the correlation of the parameters would have failed to conclude that increasing time after diagnosis influences time-conditional survival.

Our formal statistical methodology is an improvement upon the profile-based approach, where profiles are created based on the covariate patterns, as it allows for the evaluation of the statistical significance of profiles and of factors. To name a few, our methods can (1) assess whether surviving an additional 6-months after diagnosis influences the likelihood of surviving 3 more years, (2) whether a single profile fits a quadratic trend, (3) whether patients with and without nodal procedure and ulceration of the lesion have differing profiles, and (4) whether profiles are the same showing no change in time-conditional survival probabilities with additional time survived. The methodology proposed here is robust and generalizable to other malignancies and diseases. For example, we can consider either overall survival or progression-free survival and not just disease-specific survival in samples ranging in sample size from 200 when there is no censoring, and from 800 when censoring is as high as 35%. In this way, this methodology can help patients and clinicians weigh choices for surveillance for recurrent disease by adding conditional survival to risk prediction, while adjusting for covariates in the evaluation of overall survival. The methodology described here provides the clinician and the investigator with the ability to evaluate the clinical importance of profiles and factors by assessing whether there is a statistically significant relationship between time-conditional survival and additional time survived.

Table 2.1: SEER melanoma contrasts of profile-based differences

Contrast	Estimated Difference	Test Statistic	p-value†
Omnibus	–	18.365	0.0004
$\log CS_1 - \log CS_2$	0.0113	9.179	0.0025
$\log CS_2 - \log CS_3$	0.0002	0.002	0.9670
$\log CS_3 - \log CS_4$	−0.0154	9.184	0.0024

† Bonferroni adjusted significance level is 0.0167

Table 2.2: Estimates of log time-conditional melanoma-specific survival probabilities, their variance-covariance matrix, and estimates from the saturated (H_1), QM, LM, and GM models

	H_1	Estimated Matrix*				H_0 (QM†)	H_0 (LM††)	H_0 (GM††)
$\log \widehat{CS}_1$	-0.132	0.513	0.503	0.459	0.392	-0.132	-0.128	-0.129
$\log \widehat{CS}_2$	-0.143	0.883	0.632	0.589	0.521	-0.144	-0.130	-0.129
$\log \widehat{CS}_3$	-0.144	0.705	0.815	0.826	0.759	-0.143	-0.131	-0.129
$\log \widehat{CS}_4$	-0.128	0.562	0.672	0.857	0.950	-0.128	-0.133	-0.129
TS						0.14	18.01	18.36
p-value						0.7107	0.0001	0.0004

* Covariances and variances of log time-conditional survival estimates are obtained by multiplying upper and diagonal elements by a factor of 10^4, respectively. Lower diagonal elements are correlations among log time-conditional estimates

† $p > 0.05$

†† $p < 0.0001$

Table 2.3: Parameter estimates for the multivariable analysis

Estimated Parameter (β)	$\hat{\boldsymbol{\beta}}_f^\dagger$	$\hat{\boldsymbol{\beta}}_r^\dagger$	$\hat{\boldsymbol{\beta}}_A^\dagger$	$\hat{\boldsymbol{\beta}}_B^\dagger$
β_{01}	-0.148	-0.165	-0.197	-0.067
β_{02}	-0.160	-0.176	-0.209	-0.078
β_{03}	-0.162	-0.150	-0.188	-0.078
β_{04}	-0.136	-0.123	-0.152	-0.069
β_{11}	0.094	0.112	0.119	
β_{12}	0.095	0.113	0.119	
β_{13}	0.096	0.085	0.096	
β_{14}	0.062	0.048	0.062	
β_{21}	-0.134	-0.087		-0.092
β_{22}	-0.136	-0.090		-0.094
β_{23}	-0.090	-0.094		-0.092
β_{24}	-0.059	-0.069		-0.073
β_{31}	0.054			
β_{32}	0.054			
β_{33}	-0.014			
β_{34}	-0.022			

† The strata are represented by $\{X_1, X_2\} \in \{(0,0),(0,1),(1,0),(1,1)\}$ defining clinically-staged patients with no ulceration of the lesion $(k = 1)$, clinically-staged patients with an ulcerated lesion $(k = 2)$, pathologically-staged patients with no ulceration of the lesion $(k = 3)$, and pathologically-staged patients with an ulcerated lesion $(k = 4)$, respectively

Table 2.4: Estimates of time-conditional melanoma-specific survival probabilities from the multivariable analysis

Strata† (k)	Time Point (p)	Estimated CS_{kp}	$\mathbf{CS}_f$	$\mathbf{CS}_r$	$\mathbf{CS}_A$	$\mathbf{CS}_B$
1	1	CS_{11}	0.862	0.848	0.821	0.935
	2	CS_{12}	0.852	0.839	0.811	0.925
	3	CS_{13}	0.850	0.861	0.829	0.925
	4	CS_{14}	0.873	0.884	0.859	0.933
2	1	CS_{21}	0.754	0.777	0.821	0.853
	2	CS_{22}	0.744	0.767	0.811	0.842
	3	CS_{23}	0.778	0.783	0.829	0.844
	4	CS_{24}	0.823	0.825	0.859	0.868
3	1	CS_{31}	0.946	0.948	0.925	0.935
	2	CS_{32}	0.937	0.939	0.914	0.925
	3	CS_{33}	0.937	0.937	0.912	0.925
	4	CS_{34}	0.929	0.929	0.914	0.933
4	1	CS_{41}	0.875	0.869	0.925	0.853
	2	CS_{42}	0.863	0.859	0.914	0.842
	3	CS_{43}	0.845	0.853	0.912	0.844
	4	CS_{44}	0.857	0.867	0.914	0.868

† The strata are represented by $\{X_1, X_2\} \in \{(0,0),(0,1),(1,0),(1,1)\}$ defining clinically-staged patients with no ulceration of the lesion $(k = 1)$, clinically-staged patients with an ulcerated lesion $(k = 2)$, pathologically-staged patients with no ulceration of the lesion $(k = 3)$, and pathologically-staged patients with an ulcerated lesion $(k = 4)$, respectively

CHAPTER 3

PARAMETRIC TIME-CONDITIONAL SURVIVAL PROBABILITY

3.1. Introduction

Time-conditional survival probability is defined as the probability of surviving an additional Δ years beyond a, given that survival is greater than a years and can be expressed as the ratio of the a- and $(a + \Delta)$-year survival probabilities. A majority of current medical literature on time-conditional survival presents point estimates and confidence intervals using nonparametric estimates of survival time. In the nonparametric framework, numerator and denominator probabilities are typically estimated from a single Kaplan-Meier survivor function (Kaplan and Meier, 1958). These nonparametric methods stratify patients into groups based on categorical variables (requiring categorization of continuous variables) to define covariate patterns.

We present a parametric approach to estimation of time-conditional survival analysis. In Section 3.2 we derive the large sample distribution for time-conditional survival probability estimated from the Weibull survival distribution adjusting for continuous covariates and from a Logistic-Weibull cure model adjusting for continuous covariates. The hypothesis testing framework used to compare point estimates takes into account the correlation among the survival probabilities in the estimation of time-conditional survival probability. In Section 3.3, we use this approach for time-conditional survival probability estimated from the Weibull regression model and from the Logistic-Weibull cure model. This methodology is applied to esophageal cancer and melanoma data in Section 3.4, with discussion of findings in Section 3.5.

3.2. Parametric Time-Conditional Survival

We develop a general approach to derive parametric time-conditional survival probability estimators and the approximate large sample distribution for these estimators. To estimate time-conditional survival adjusting for continuous covariates, we provide an approach using a parametric regression model and an approach using a cure model. Sections 3.2.1 and 3.2.2 define time-conditional survival probability as a function of the maximum likelihood estimators of the model parameters and covariate coefficients. Section 3.3 provides greater detail on the time-conditional survival probability

maximum likelihood estimator under the Weibull survival model and the logistic-Weibull cure model.

3.2.1. Continuous Covariate-Adjusted Time-Conditional Survival Estimation

A strength of the parametric approach to time-conditional survival probability estimation is the ability to incorporate information from multiple continuous covariates.

Let T denote the time to the event of interest. Then, let the survival data be captured for a sample of n patients such that the ith individual contributes observation $(t_i, \delta_i), i = 1, \ldots, n$. The indicator function δ_i is set to unity when the observed survival time, t_i, is an event time, and to zero when it is a censored time.

The partial likelihood function is given by

$$\prod_{i=1}^{n} (f(t_i \mid \boldsymbol{\theta}))^{\delta_i} (S(t_i \mid \boldsymbol{\theta}))^{1-\delta_i},$$

where $f(\cdot)$ is a probability density function and $S(\cdot)$ is a survivor function. These are written as functions of a vector of distribution parameters, $\boldsymbol{\theta}$, which generally represents more than one parameter. The approximate large sample joint distribution of the maximum likelihood estimators, $\hat{\boldsymbol{\theta}}$, of the parameters from the survival distribution is given by

$$\hat{\boldsymbol{\theta}} \xrightarrow{d} N(\boldsymbol{\theta}, \boldsymbol{\Sigma}_\theta),$$

where $\boldsymbol{\Sigma}_\theta$ is estimated by the inverse of the observed Fisher information (Efron and Hinkley, 1978).

Let $\mathbf{Z}$ denote a vector of continuous covariates. The accelerated failure time model is defined by

$$S(t \mid \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\theta}) = S_0(t \exp(\boldsymbol{\beta}'\mathbf{Z}) \mid \boldsymbol{\theta}), \tag{3.1}$$

for all values of time t, where $\boldsymbol{\beta}$ is a vector of regression coefficients. Alternatively, assuming a linear relationship between log time and the continuous covariates, the linear model is given by

$$Y = \log T = \mu + \boldsymbol{\gamma}'\mathbf{Z} + \sigma W, \tag{3.2}$$

where $\boldsymbol{\gamma}$ is a vector of regression parameters. The distribution of Y is determined by the (error)

distribution given by W, or the choice of S_0 in Equation 3.1 under the accelerated failure time model.

The parametric covariate-adjusted time-conditional survival probability is expressed as transformations of the k parameters $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ and is given by

$$CS(a+\Delta \mid a, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\theta}) = \frac{S(a+\Delta \mid \boldsymbol{\beta}, \boldsymbol{\theta})}{S(a \mid \boldsymbol{\beta}, \boldsymbol{\theta})}.$$

This represents the probability of surviving an additional Δ time units beyond a, given that survival is greater than time a. By the invariance property, the maximum likelihood estimator of the covariate-adjusted time-conditional survival probability is given by

$$\widehat{CS}(a+\Delta \mid a, \mathbf{Z}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) = \frac{\hat{S}((a+\Delta)\exp(\hat{\boldsymbol{\beta}}'\mathbf{Z}) \mid \hat{\boldsymbol{\theta}})}{\hat{S}(a\exp(\hat{\boldsymbol{\beta}}'\mathbf{Z}) \mid \hat{\boldsymbol{\theta}})}.$$

By the δ-method, the large sample expectation and variance of this transformation are given by

$$E\left(\widehat{CS}(a+\Delta \mid a, \mathbf{Z}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})\right) = CS(a+\Delta \mid a, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\theta}),$$

and

$$Var\left(\widehat{CS}(a+\Delta \mid a, \mathbf{Z}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})\right) = J(\boldsymbol{\beta}, \boldsymbol{\theta})^T \boldsymbol{\Sigma}_{\beta,\theta} J(\boldsymbol{\beta}, \boldsymbol{\theta}),$$

respectively, where $J(\boldsymbol{\beta}, \boldsymbol{\theta})$ is the Jacobian $k \times 1$ vector of first degree partial derivatives of time-conditional survival probabilities,with respect to the k parameters, $(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})^T$, and is given by

$$J(\boldsymbol{\beta}, \boldsymbol{\theta}) = \begin{pmatrix} \frac{\partial CS}{\partial \beta_1} \\ \frac{\partial CS}{\partial \beta_2} \\ \vdots \\ \frac{\partial CS}{\partial \beta_{c_1}} \\ \frac{\partial CS}{\partial \theta_1} \\ \vdots \\ \frac{\partial CS}{\partial \theta_{c_2}} \end{pmatrix},$$

where $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_{c_1})^T$, $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{c_2})^T$ and where $c_1 + c_2 = k$. By the multivariate δ-method

and Slutsky's Theorem, the variance is estimated by

$$\widehat{Var}\left(\widehat{CS}(a+\Delta \mid a, \mathbf{Z}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})\right) = \widehat{J}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})^T \widehat{\boldsymbol{\Sigma}}_{\hat{\beta},\hat{\theta}} \widehat{J}(\hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}})$$

and for the Weibull survival distribution, the partial derivatives are derived in the Appendix. The large sample distribution of the maximum likelihood estimator of parametric time-conditional survival probability is given by

$$\widehat{CS}(a+\Delta \mid a, \mathbf{Z}, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\theta}}) \xrightarrow{d} N\left(CS(a+\Delta \mid a, \mathbf{Z}, \boldsymbol{\beta}, \boldsymbol{\theta}), J(\boldsymbol{\beta}, \boldsymbol{\theta})^T \boldsymbol{\Sigma}_{\beta,\theta} J(\boldsymbol{\beta}, \boldsymbol{\theta})\right).$$

3.2.2. Cure Model Based Time-Conditional Survival Estimation

The cure model assumes that the underlying population is a mixture of subjects who will or will not experience the event, where cured subjects will not experience the event. As with the parametric regression model methods described in Section 3.2.1, the cure model can incorporate continuous and categorical variables. A method for time-conditional survival probability analysis based on a cure model adjusting for covariates follows.

Define T to be the time to the event of interest and let t be the observed time. Define E to be an indicator of cure status such that $E = 1$ represents those individuals not cured of the event and $E = 0$ represents those individuals cured of the event. The random variables T and E each follow a parametric distribution. The mixture cure model is given by

$$S(t \mid \boldsymbol{\zeta}, \boldsymbol{\eta}, \mathbf{x}, \mathbf{z}) = \pi(\boldsymbol{\eta}, \mathbf{z}) \times S_e(t \mid E = 1, \boldsymbol{\zeta}, \mathbf{x}) + 1 - \pi(\boldsymbol{\eta}, \mathbf{z}), \tag{3.3}$$

where the continuous covariate vectors $\mathbf{x}$ and $\mathbf{z}$ may be the same or may differ. Here $S_e(t \mid E = 1, \boldsymbol{\zeta}, \mathbf{x}) = P(T > t \mid E = 1, \boldsymbol{\zeta}, \mathbf{x})$ is the survival probability for individuals not cured of the event given a vector of continuous covariates $\mathbf{x}$ and given a vector of regression parameters $\boldsymbol{\zeta}$. The survival distribution for individuals not cured of the event is generally modeled using parametric or semi-parametric survival models. The probability of observing an individual not cured is $\pi(\boldsymbol{\eta}, \mathbf{z}) = P(E = 1 \mid \boldsymbol{\eta}, \mathbf{z})$ where $\mathbf{z}$ is the continuous covariate vector and where $\boldsymbol{\eta}$ is a function of regression parameters. The influence of the continuous covariates on the probability of no cure is modeled using a logistic regression model.

The observed data take the form $(t_i, \delta_i, \mathbf{x}_i, \mathbf{z}_i)$ for $i = 1, \ldots, n$. Here, t_i is the observed time on study and δ_i is the censoring indicator. For individual i, when $\delta_i = 1$, the contribution to the likelihood is given by

$$\pi_i(\boldsymbol{\eta}, \mathbf{z}_i) f_e(t_i \mid E = 1, \mathbf{x}_i),$$

where $f_e(\cdot)$ is the probability density function for time to an event for individuals not cured. When $\delta_i = 0$, the contribution to the likelihood is given by

$$(1 - \pi_i(\boldsymbol{\eta}, \mathbf{z}_i)) + \pi_i(\boldsymbol{\eta}, \mathbf{z}_i) S_e(t_i \mid E = 1, \boldsymbol{\zeta}, \mathbf{x}_i).$$

Then the observed likelihood is given by

$$\begin{aligned} L(\boldsymbol{\eta}, \boldsymbol{\zeta}) = \prod_{i=1}^{n} & \left(\pi_i(\boldsymbol{\eta}, \mathbf{z}_i) f_e(t_i \mid E = 1, \boldsymbol{\zeta}, \mathbf{x}_i)\right)^{\delta_i} \\ & \times \left((1 - \pi_i(\boldsymbol{\eta}, \mathbf{z}_i)) + \pi_i(\boldsymbol{\eta}, \mathbf{z}_i) S_e(t_i \mid E = 1, \boldsymbol{\zeta}, \mathbf{x}_i)\right)^{1-\delta_i}. \end{aligned}$$

Under this framework, time-conditional survival probability is defined as

$$\begin{aligned} CS(a + \Delta \mid a, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\zeta}, \mathbf{x}, \mathbf{z}) &= \frac{S(a + \Delta)}{S(a)} \\ &= \frac{(1 - \pi(\boldsymbol{\eta}, \mathbf{z})) + \pi(\boldsymbol{\eta}, \mathbf{z}) S_e(a + \Delta \mid E = 1, \boldsymbol{\zeta}, \mathbf{x})}{(1 - \pi(\boldsymbol{\eta}, \mathbf{z})) + \pi(\boldsymbol{\eta}, \mathbf{z}) S_e(a \mid E = 1, \boldsymbol{\zeta}, \mathbf{x})}, \end{aligned}$$

where $\boldsymbol{\eta}$ and $\boldsymbol{\zeta}$ correspond to the regression parameters associated with the continuous covariate vectors $\mathbf{z}$ and $\mathbf{x}$ and the survival distribution parameters $\boldsymbol{\beta}$ and $\boldsymbol{\gamma}$, respectively. The large sample distribution for an estimator of time-conditional survival probability is given by

$$\widehat{CS}(a + \Delta \mid a, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\zeta}}, \mathbf{x}, \mathbf{z}) \xrightarrow{d} N\left(CS(a + \Delta \mid a, \boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\zeta}, \mathbf{x}, \mathbf{z}), \sigma^2_{CS}\right).$$

The large sample variance of time-conditional survival probability, σ^2_{CS} is given by $J(\boldsymbol{\theta})^T \Sigma_\theta J(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\eta}, \boldsymbol{\zeta})^T$, where $J(\boldsymbol{\theta})$ is a vector of first degree partial derivatives given by

$$J(\boldsymbol{\theta}) = \left(\left(\frac{\partial CS}{\partial \boldsymbol{\beta}}\right)^T, \left(\frac{\partial CS}{\partial \boldsymbol{\gamma}}\right)^T, \left(\frac{\partial CS}{\partial \boldsymbol{\eta}}\right)^T, \left(\frac{\partial CS}{\partial \boldsymbol{\zeta}}\right)^T\right)^T.$$

By the multivariate δ-method and Slutsky's Theorem, the variance is estimated by

$$\widehat{Var}\left(\widehat{CS}(a+\Delta \mid a, \hat{\boldsymbol{\beta}}, \hat{\boldsymbol{\gamma}}, \hat{\boldsymbol{\eta}}, \hat{\boldsymbol{\zeta}}, \mathbf{x}, \mathbf{z})\right) = \widehat{J}(\hat{\boldsymbol{\theta}})^T \widehat{\boldsymbol{\Sigma}}_{\hat{\theta}} \widehat{J}(\hat{\boldsymbol{\theta}})$$

and for the Logistic-Weibull cure model, the partial derivatives are derived in the Appendix.

3.2.3. Comparing Across Profiles

After a time-conditional survival probability function is defined, the maximum likelihood estimators will be obtained and used to make inferences. For example, fixing some of the continuous covariates and varying others can produce a series of profiles. A relevant question is whether these profiles, which are estimated from a single sample using maximum likelihood theory, are significantly different with respect to time-conditional survival. We can address this question by fixing the time survived after diagnosis and comparing the profiles at a fixed point in time, as shown below.

Clinical studies of time-conditional survival have used a profile of estimated time-conditional survival probabilities (shown here as a $p \times 1$ vector) given by

$$\widehat{\mathbf{CS}} = \left(\widehat{CS}_1(b_1 \mid a_1), \widehat{CS}_2(b_2 \mid a_2), \ldots, \widehat{CS}_p(b_p \mid a_p)\right)^T.$$

When $b_j = a_j + \Delta$ for $j = 1, \ldots, p$, these estimators represent consecutive Δ-year time-conditional survival probabilities. For example, 5-year time-conditional survival probabilities, given that survival is greater than 1, 2, and 3 years after diagnosis, are consecutive estimators that can be expressed as

$$\widehat{\mathbf{CS}}_{3\times 1} = \left(\widehat{CS}_1(6 \mid 1), \widehat{CS}_2(7 \mid 2), \widehat{CS}_3(8 \mid 3)\right)^T.$$

To develop the methods in this section, it is helpful to keep in mind the following example. Suppose there is a disease registry with patient-level data on age at diagnosis (continuous), sex (males/females), and three stages of the disease (denoted as Stage I, II, or III). This section discusses estimation and hypothesis testing for 5-year time-conditional survival probability as a function of time survived after diagnosis (denoted as a) and covariates.

Comparing two point estimates: an example. Suppose we are interested in whether the probability of surviving an additional 5 years differs for a 45 year old patient as compared to a 65 year old

patient. To address this question we first obtain the maximum likelihood estimates for the parametric survival function. To obtain a point estimate of time-conditional survival, we fix sex and stage at specific values and fix $\Delta = 5$.

By fixing time and covariates, point estimates for time-conditional survival probabilities obtained from different profiles can be compared when $a = 1$. In this scenario, the 5-year time-conditional survival probability, given that survival is greater than 1 year after diagnosis, is estimated for the 45 year old patients and for the 65 year old patients. The null hypothesis of no difference between the two probabilities and the alternative hypothesis are given by

$$H_0 : CS_{ia} - CS_{ja} = 0 \qquad \text{and} \qquad H_1 : CS_{ia} - CS_{ja} \neq 0. \tag{3.4}$$

Let the index i represent the 45 year old patients' profiles, the index j represent the 65 year old patients' profiles, and let a represent the time alive after diagnosis. After fitting the survival model to the data, we obtained estimates of the regression parameters. In order to visualize the results and to conduct this hypothesis test, point estimates and their variances must be obtained by plugging in certain values for the continuous covariate, time survived, and additional time survived.

Define the Wald χ^2 test statistic given by

$$TS(CS_{ia} - CS_{ja} = 0) = \frac{\left(\widehat{CS}_{ia} - \widehat{CS}_{ja}\right)^2}{\widehat{Var}(\widehat{CS}_{ia} - \widehat{CS}_{ja})} \sim \chi^2_{(1,\alpha=0.05)}, \tag{3.5}$$

where the estimated variance is given by

$$\widehat{Var}(\widehat{CS}_{ia} - \widehat{CS}_{ja}) = \widehat{Var}(\widehat{CS}_{ia}) + \widehat{Var}(\widehat{CS}_{ja}) - 2 \cdot \widehat{Cov}(\widehat{CS}_{ia}, \widehat{CS}_{ja}).$$

This statistic allows the researcher to determine whether the two time-conditional survival estimates are significantly different. For example, from the parametric regression model, the researcher can evaluate whether 5-year time-conditional survival, given that survival is greater than 1 year, is significantly different for a 45 year old male with stage II disease as compared to a 65 year old male with stage I disease.

Comparing K point estimates. To describe the comparison of K point estimates, we consider

tumor length as an example. Because patients are concerned about whether a tumor will shorten their life (overall survival) as well as their future quality of life, their probability of future survival may impact choices about timing and invasiveness of treatment. Tumor length may be the most influential prognostic variable at diagnosis (Vollmer, 2008). Time-conditional survival probability estimates from parametric models can be computed and compared for specific values of tumor length to assess the impact of this continuous covariate.

Consider three possible profiles in which time survived after diagnosis may vary by tumor length. Our methodology can be used to assess whether the probability of surviving an additional 5 years varies for several patient-specific tumor lengths such as 1 mm, 2 mm, and 3 mm. In order to draw these comparisons, we first obtain the estimates of the parameters for the survival distribution and then estimates of time-conditional survival probability with fixed time survived (a) and other covariates.

Generalizing to K groups ($k = 1, \ldots, K$), the null hypothesis is

$$H_0 : CS_{1a} = CS_{2a} = \ldots = CS_{Ka}.$$

In this example, the 3 groups ($K = 3$) are defined by tumor length (1 mm, 2 mm, and 3 mm), and a represents the time alive after diagnosis, which is consistent with the notation for a in the estimation of time-conditional survival probability. The null hypothesis in matrix notation is given by

$$H_0 : \begin{pmatrix} 1 & -1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \ldots & -1 \end{pmatrix} \begin{pmatrix} CS_{1a} \\ CS_{2a} \\ \vdots \\ CS_{Ka} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{3.6}$$

where the design matrix, $\mathbf{X}$, is $(k-1) \times k$ and the time-conditional survival probability vector has length k. In words is a test of whether these differences are all zero and, therefore, whether

$CS_{1a}, CS_{2a}, \ldots, CS_{Ka}$ are all equal. The alternative hypothesis is given by

$$H_1 : \begin{pmatrix} 1 & -1 & \ldots & 0 \\ \vdots & \vdots & & \vdots \\ 1 & 0 & \ldots & -1 \end{pmatrix} \begin{pmatrix} CS_{1a} \\ CS_{2a} \\ \vdots \\ CS_{Ka} \end{pmatrix} = \begin{pmatrix} \delta_1 \\ \delta_2 \\ \vdots \\ \delta_{k-1} \end{pmatrix}, \tag{3.7}$$

i.e., at least one of the differences specified under the null hypothesis is not equal to zero and, therefore, $CS_{1a}, CS_{2a}, \ldots, CS_{Ka}$ are not all equal.

$K - 1$ comparisons are created when using one group as the reference among the K groups. The test statistic is given by

$$\left(\mathbf{X}\widehat{\mathbf{CS}}\right)' \left(\mathbf{X}\hat{\mathbf{\Sigma}}\mathbf{X}'\right)^{-1} \left(\mathbf{X}\widehat{\mathbf{CS}}\right) \sim \chi^2_{(\textit{Rank(X)},\alpha=0.05)} \tag{3.8}$$

where *Rank(X)* represents the rank of $\mathbf{X}$. For the example with three tumor lengths of interest, the degrees of freedom are *Rank(X)* $= 2$. The elements of the estimated covariance matrix, $\hat{\mathbf{\Sigma}}$, are derived in the Appendix and their functional form varies depending on whether the probabilities are estimated from the covariate-adjusted survival model or the cure model.

A similar hypothesis testing framework can be applied to compare multiple profiles at a given point in time for a covariate with multiple levels such as disease stage. Consider three profiles based on disease stage in which time survived after diagnosis may vary. For this scenario, where stage I is the referent group, we can assess whether the probability of surviving an additional 5 years varies for patients with stage II disease as compared to those with stage I disease and, similarly, whether the probability of surviving an additional 5 years varies for patients with stage III disease as compared to those with stage I disease.

Contrasts between groups. Consider the probability of surviving an additional 5 years for patients aged 45 versus 65 years old across the three tumor lengths (1, 2, and 3 mm). The analysis would focus on six estimates of the same time-conditional survival probability model for patients in six different groups (three tumor lengths for each of the two ages of interest) and, for a fixed value of time survived from diagnosis, $a = 1$ using three comparisons, using the differences between the time-conditional survival probability patients 45 years compared to 65 years for each level of tumor

thickness.

The design matrix specifies the comparison of time-conditional survival probability across these six estimators and the null hypothesis is given by

$$H_0: \begin{array}{l} CS_{111} - CS_{211} = 0 \\ CS_{121} - CS_{221} = 0 \\ CS_{131} - CS_{231} = 0 \end{array}, \tag{3.9}$$

and is written in matrix notation as

$$H_0: \begin{pmatrix} 1 & 0 & 0 & -1 & 0 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} CS_{111} \\ CS_{121} \\ CS_{131} \\ CS_{211} \\ CS_{221} \\ CS_{231} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}. \tag{3.10}$$

In this hypothesis test, the indices for CS_{ija} represent patients aged 45 ($i = 1$) or 65 ($i = 2$) years old, j represents tumor length ($j = 1, 2, 3$ for tumor lengths 1 mm, 2 mm, and 3 mm, respectively), and a represents time survived from diagnosis. This null hypothesis states that there is no difference between 45 and 65 year old patients with the same tumor length of 1, 2, or 3 mm. By estimating 5-year time-conditional survival probability, given that survival is greater than 1 year after diagnosis, it is possible to determine if, for at least one value of tumor length, there is a significant difference in time-conditional survival probability for 45 year old patients as compared to 65 year old patients.

Comparing profiles of estimators over discrete time. Similar methods are used to compare time-conditional survival probabilities across fixed times after diagnosis, a, for different patient profiles. For example, the 5-year profiles of time-conditional survival probability estimates for 45 and 65 year old patients can be compared using the following vector of time-conditional survival probabilities

$$\widehat{\mathbf{CS}} = \left(\widehat{CS}(1+5 \mid 1, \hat{\boldsymbol{\theta}}), \widehat{CS}(2+5 \mid 2, \hat{\boldsymbol{\theta}}), \widehat{CS}(3+5 \mid 3, \hat{\boldsymbol{\theta}})\right)^T,$$

which denotes the 5-year time-conditional survival probabilities, given that survival is greater than

1, 2, and 3 years after diagnosis, for fixed covariate values.

Under this framework, to compare profiles over discrete time, Equations 3.9 and 3.10 refer to the null hypothesis that the estimates of time-conditional survival probabilities for 45 and 65 year old patients are the same across these discrete time points where the indices for CS_{ij} represent 45 ($i = 1$) or 65 ($i = 2$) year old patients, but j now represents time alive after diagnosis in years ($j = 1, 2, 3$). The alternative hypothesis is that at least one of these differences is not equal to zero, meaning that the two profiles are different. The test statistic has the same functional form as in Equation 3.8 and is distributed as χ^2 with *Rank(X)* degrees of freedom and $\alpha = 0.05$. We illustrate this approach in Section 3.4.

3.3. An Example: The Weibull Distribution

To derive and demonstrate some of the relationships in the section above, let T have a Weibull distribution. Then the continuous covariate-adjusted time-conditional survival estimation and the Logistic-Weibull cure model based time-conditional survival estimation can be obtained.

3.3.1. Adjusted Weibull Time-Conditional Survival

Consider a vector of continuous covariates, $\mathbf{z}$, and define the vector of survival distribution parameters as $\boldsymbol{\theta}$. We fit a log linear survival model for log time, adjusting for continuous covariates as given by Equation 3.2 where W follows the extreme value distribution. (Estimates of the extreme value distribution parameters are found numerically and exist in statistical software packages.) The survival function of $\log T$ adjusting for the covariates, $\mathbf{z}$, is given by

$$S(\log t \mid \mathbf{z}, \mu, \sigma, \boldsymbol{\gamma}) = \exp\left(-\exp\left(\frac{\log t - \mu - \boldsymbol{\gamma}'\mathbf{z}}{\sigma}\right)\right),$$

where γ represents the vector of regression coefficients on the $\log$ scale and μ and σ represent the survival distribution parameters. The survival function can be written as a model for T when $\alpha = 1/\sigma, \lambda = \exp(-\mu/\sigma)$, and $\beta_j = \gamma_j/\sigma$ and is given by

$$S(t \mid \mathbf{z}, \alpha, \lambda, \boldsymbol{\beta}) = \exp\left(-t^{\alpha} \cdot \lambda \cdot \exp\left(\boldsymbol{\beta}'\mathbf{z}\right)\right),$$

(Klein and Moeschberger, 2005).

We obtain maximum likelihood estimates of the parameters under the Weibull parametric model by estimating the extreme value distribution model parameters and their covariance matrix given by

$$\hat{\alpha} = 1/\hat{\sigma}, \qquad \hat{\lambda} = \exp(-\hat{\mu}/\hat{\sigma}), \qquad \hat{\beta}_j = -\hat{\gamma}_j/\hat{\sigma}, \qquad j = 1, \ldots, p$$

$$Var(\hat{\alpha}) = \frac{Var(\hat{\sigma})}{\sigma^4},$$

$$Var(\hat{\lambda}) = \exp\left(-\frac{2\mu}{\sigma}\right) \times \left(\frac{Var(\hat{\mu})}{\sigma^2} - \frac{2\mu Cov(\hat{\mu}, \hat{\sigma})}{\sigma^3}\right) + \frac{\mu^2 Var(\hat{\sigma})}{\sigma^4},$$

$$Cov(\hat{\alpha}, \hat{\lambda}) = \exp\left(\frac{\mu}{\sigma}\right) \times \left(\frac{Cov(\hat{\mu}, \hat{\sigma})}{\sigma^3} - \frac{\mu Var(\hat{\sigma})}{\sigma^4}\right),$$

and for $l, m = 1, \ldots, k$,

$$Cov(\hat{\beta}_l, \hat{\beta}_m) = \frac{Cov(\hat{\gamma}_l, \hat{\gamma}_m)}{\sigma^2} - \frac{\gamma_l Cov(\hat{\gamma}_l, \hat{\sigma})}{\sigma^3} - \frac{\gamma_m Cov(\hat{\gamma}_m, \hat{\sigma})}{\sigma^3} + \frac{\gamma_l \gamma_m Var(\hat{\sigma})}{\sigma^4},$$

$$Cov(\hat{\beta}_l, \hat{\alpha}) = \frac{Cov(\hat{\gamma}_l, \hat{\sigma})}{\sigma^3} - \frac{\gamma_l Var(\hat{\sigma})}{\sigma^4},$$

$$Cov(\hat{\beta}_l, \hat{\lambda}) = \exp\left(\frac{\mu}{\sigma}\right) \times \left(\frac{Cov(\hat{\gamma}_l, \hat{\mu})}{\sigma^2} - \frac{\gamma_l Cov(\hat{\gamma}_l, \hat{\sigma})}{\sigma^3} - \frac{\mu Cov(\hat{\mu}, \hat{\sigma})}{\sigma^3}\right) + \frac{\gamma_l \mu Var(\hat{\sigma})}{\sigma^4}.$$

Define the vector of parameters as given by $\boldsymbol{\theta}^* = (\mu, \sigma, \gamma_1, \ldots, \gamma_p)^T$. Then, the large sample joint distribution is given by

$$\hat{\boldsymbol{\theta}}^* = \begin{pmatrix} \hat{\mu} \\ \hat{\sigma} \\ \hat{\gamma}_1 \\ \vdots \\ \hat{\gamma}_p \end{pmatrix} \xrightarrow{d} N \left(\boldsymbol{\theta}^* = \begin{pmatrix} \mu \\ \sigma \\ \gamma_1 \\ \vdots \\ \gamma_p \end{pmatrix}, \ \boldsymbol{\Sigma}_{\theta^*} \right),$$

where

$$\boldsymbol{\Sigma}_{\theta^*} = \begin{pmatrix} Var(\hat{\mu}) & Cov(\hat{\mu}, \hat{\sigma}) & Cov(\hat{\mu}, \hat{\gamma}_1) & \cdots & Cov(\hat{\mu}, \hat{\gamma}_p) \\ Cov(\hat{\mu}, \hat{\sigma}) & Var(\hat{\sigma}) & Cov(\hat{\sigma}, \hat{\gamma}_1) & \cdots & Cov(\hat{\sigma}, \hat{\gamma}_p) \\ Cov(\hat{\mu}, \hat{\gamma}_1) & Cov(\hat{\sigma}, \hat{\gamma}_1) & Var(\hat{\gamma}_1) & \cdots & Cov(\hat{\gamma}_1, \hat{\gamma}_p) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ Cov(\hat{\mu}, \hat{\gamma}_p) & Cov(\hat{\sigma}, \hat{\gamma}_p) & Cov(\hat{\gamma}_1, \hat{\gamma}_p) & \cdots & Var(\hat{\gamma}_p) \end{pmatrix},$$

as $n \to \infty$, such that $\boldsymbol{\Sigma}_{\theta^*}$ is the inverse of the Fisher information. The maximum likelihood estimator of time-conditional survival probability given continuous covariates $\mathbf{z}$ is given by

$$\begin{aligned}\widehat{CS}(a+\Delta \mid a, \mathbf{z}, \hat{\alpha}, \hat{\lambda}, \hat{\boldsymbol{\beta}}) &= \frac{\exp\left(-(a+\Delta)^{\hat{\alpha}} \cdot \hat{\lambda} \cdot \exp\left(\hat{\boldsymbol{\beta}}'\mathbf{z}\right)\right)}{\exp\left(-a^{\hat{\alpha}} \cdot \hat{\lambda} \cdot \exp\left(\hat{\boldsymbol{\beta}}'\mathbf{z}\right)\right)} \\ &= \exp\left(\hat{\lambda} \cdot \exp\left(\hat{\boldsymbol{\beta}}'\mathbf{z}\right) \cdot \left(a^{\hat{\alpha}} - (a+\Delta)^{\hat{\alpha}}\right)\right).\end{aligned}$$

Covariance of two estimators. Define a Weibull time-conditional survival probability with a single continuous covariate, z, in the model. To investigate whether a different value of the continuous covariate has a significant influence on the time-conditional survival, consider an estimator with fixed a and Δ. The covariance between time-conditional survival probabilities must be accounted for in order to evaluate the relationship.

Define two estimators of time-conditional survival probability from a single sample at different values of the continuous covariate, $CS_i = CS(b_i \mid a_i, z_i, \boldsymbol{\theta})$ where $i = 1, 2$ and $\boldsymbol{\theta} = (\alpha, \lambda, \beta)^T$, such that

$$CS_i = \frac{\exp\left(-b_i^{\alpha}\lambda \exp(\beta \cdot z_i)\right)}{\exp\left(-a_i^{\alpha}\lambda \exp(\beta \cdot z_i)\right)}, \tag{3.11}$$

and $b_i = a_i + \Delta$. The large sample variance of time-conditional survival based on the Weibull regression model is given by

$$\begin{aligned}Var(CS_i) = Var(\hat{\alpha})\left(\frac{\partial CS_i}{\partial \alpha}\right)^2 + Var(\hat{\lambda})\left(\frac{\partial CS_i}{\partial \lambda}\right)^2 + Var(\hat{\beta})\left(\frac{\partial CS_i}{\partial \beta}\right)^2 \\ + 2\left(\frac{\partial CS_i}{\partial \alpha}\frac{\partial CS_i}{\partial \lambda}Cov(\hat{\alpha}, \hat{\lambda}) + \frac{\partial CS_i}{\partial \alpha}\frac{\partial CS_i}{\partial \beta}Cov(\hat{\alpha}, \hat{\beta}) + \frac{\partial CS_i}{\partial \lambda}\frac{\partial CS_i}{\partial \beta}Cov(\hat{\lambda}, \hat{\beta})\right).\end{aligned}$$

For any two time-conditional survival estimators obtained from this parametrization of the Weibull regression model, the large sample covariance is given by

$$\begin{aligned}Cov(\widehat{CS}_1, \widehat{CS}_2) = &\frac{\partial CS_1}{\partial \alpha}\frac{\partial CS_2}{\partial \alpha}Var(\hat{\alpha}) + \frac{\partial CS_1}{\partial \lambda}\frac{\partial CS_2}{\partial \alpha}Cov(\hat{\alpha}, \hat{\lambda}) \\ &+ \frac{\partial CS_1}{\partial \beta}\frac{\partial CS_2}{\partial \alpha}Cov(\hat{\alpha}, \hat{\beta}) + \frac{\partial CS_1}{\partial \alpha}\frac{\partial CS_2}{\partial \lambda}Cov(\hat{\alpha}, \hat{\lambda}) \\ &+ \frac{\partial CS_1}{\partial \lambda}\frac{\partial CS_2}{\partial \lambda}Var(\hat{\lambda}) + \frac{\partial CS_1}{\partial \beta}\frac{\partial CS_2}{\partial \lambda}Cov(\hat{\lambda}, \hat{\beta}) \\ &+ \frac{\partial CS_1}{\partial \alpha}\frac{\partial CS_2}{\partial \beta}Cov(\hat{\alpha}, \hat{\beta}) + \frac{\partial CS_1}{\partial \lambda}\frac{\partial CS_2}{\partial \beta}Cov(\hat{\lambda}, \hat{\beta}) \\ &+ \frac{\partial CS_1}{\partial \beta}\frac{\partial CS_2}{\partial \beta}Var(\hat{\beta}).\end{aligned}$$

Substituting the partial derivatives from Section B.1 of the Appendix, we obtain the estimator of the large sample covariance of any two time-conditional survival probabilities based on this parameterization of the Weibull distribution. The test statistic for the null hypothesis of no difference between the estimators of time-conditional survival probabilities is given in Equation 3.8 and is distributed as χ^2 with degrees of freedom equal to *Rank(X)* and $\alpha = 0.05$.

3.3.2. Logistic-Weibull Cure Model for Time-Conditional Survival

To estimate time-conditional survival probabilities using the Logistic-Weibull cure model, we first obtain estimates of the parameters in the cure model. The probability of not being cured of the event can be modeled using a binary regression model with a logit link given by

$$logit\left(\pi(\boldsymbol{\eta}, \mathbf{z})\right) = \boldsymbol{\eta}'\mathbf{z},$$

where $\mathbf{z}$ represents a vector of continuous covariates. The survival distribution for individuals who are not cured can be modeled using the Weibull parametric regression model given by

$$S(t \mid E = 1, \boldsymbol{\zeta}, \mathbf{x}) = \exp\left(-\exp\left(\frac{\log t - \boldsymbol{\zeta}'\mathbf{x}}{\sigma}\right)\right),$$

where $\boldsymbol{\zeta}$ represents a vector of regression parameters. Using this parameterization of the mixture cure model, we obtain the parameter estimates of the continuous covariates affecting the *proportion* of not cured patients ($\mathbf{z}$) and those affecting the *survival distribution* for not cured patients ($\mathbf{x}$).

For this cure model, the survivor function is given in Equation 3.3 where $\mathbf{x}, \mathbf{z}$ may represent vectors of the same or different covariates. The functional form of these components is given by

$$\pi(\boldsymbol{\eta}, \mathbf{z}) = \frac{\exp(\boldsymbol{\eta}'\mathbf{z})}{1 + \exp(\boldsymbol{\eta}'\mathbf{z})},$$

for the logit link and

$$S_e(t \mid E = 1, \boldsymbol{\zeta}, \mathbf{x}) = \exp\left(-\exp(-\boldsymbol{\zeta}'\mathbf{x}/\sigma)t^{1/\sigma}\right),$$

for the Weibull survival distribution. Let $\alpha = 1/\sigma$ and $\lambda = \exp(-1/\sigma)$. Then the survivor function

incorporating continuous covariates is given by

$$S(t \mid \alpha, \lambda, \boldsymbol{\eta}, \boldsymbol{\zeta}, \mathbf{z}, \mathbf{x}) = \frac{\exp(\boldsymbol{\eta}'\mathbf{z}) \exp\left(-t^{\alpha}\lambda \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) + 1}{\exp(\boldsymbol{\eta}'\mathbf{z}) + 1}. \tag{3.12}$$

Adjusting for the continuous covariates, the time-conditional survival probability is given by

$$CS(b \mid a, \alpha, \lambda, \boldsymbol{\eta}, \boldsymbol{\zeta}, \mathbf{z}, \mathbf{x}) = \frac{\exp(\boldsymbol{\eta}'\mathbf{z}) \exp\left(-b^{\alpha}\lambda \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) + 1}{\exp(\boldsymbol{\eta}'\mathbf{z}) \exp\left(-a^{\alpha}\lambda \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) + 1}. \tag{3.13}$$

Consider a single continuous covariate, z, that affects the proportion of individuals who are not cured and a single continuous covariate, x, that affects the survival distribution of individuals who are not cured. The time-conditional survival probability adjusting for the continuous covariates is given by

$$CS(b \mid a, \alpha, \lambda, \eta, \zeta, z, x) = \frac{\exp(\eta_0 + \eta_1 z) \exp\left(-b^{\alpha}\lambda \exp\left(\zeta x\right)\right) + 1}{\exp(\eta_0 + \eta_1 z) \exp\left(-a^{\alpha}\lambda \exp\left(\zeta x\right)\right) + 1}.$$

We derive the form of the partial derivatives to estimate the large sample variance in Section B.2 of the Appendix.

3.4. Application to Real-World Data

To demonstrate inclusion of a continuous risk factor in the model, we applied the parametric time-conditional survival methodology to data from patients with esophageal cancer and with melanoma. Using the National Cancer Institute's Surveillance, Epidemiology, and End Results (SEER) registry, we calculated survival time from the date of diagnosis to the date of last known follow-up or death. Estimates of time-conditional disease-specific survival probabilities were computed. The author developed SAS/IML macros for the analysis of time-conditional survival probabilities (SAS Institute Inc., 2008) and used the macro $\%\mathrm{PSPMCM}$ developed by Corbière and Joly, 2007 for the estimation of cure models.

3.4.1. Adjusted Time-Conditional Survival Estimation for Esophageal Cancer

Background. Cancers of the stomach, the small intestine, and the esophagus account for roughly 3% of the annual cancer diagnoses in the United States. These cancers of the upper gastrointestinal tract have low survival rates and comprise an estimated 4.7% of the cancer deaths in the United States annually (American Cancer Society, 2006). Cancers with low survival rates can be

evaluated using parametric modeling (for example, exponential or Weibull distributions) because the data meet the assumption that, as time increases, the probability of survival tends to zero.

Sample. A cohort of 9011 patients diagnosed with a primary invasive tumor of the esophagus (ICD-O-2/3 site codes in C15.0–15.9) between 1988 and 2008 was identified from 9 SEER regions within the United States (SEER, 2008). Disease-specific survival time was defined as time to death due to esophageal cancer. In Figure 3.1, the empirical survival curve of time to death due to esophageal cancer is overlain with the estimated unadjusted Weibull survival function given the maximum likelihood estimates. This figure shows decreasing survival probabilities in the first 15 years after diagnosis.

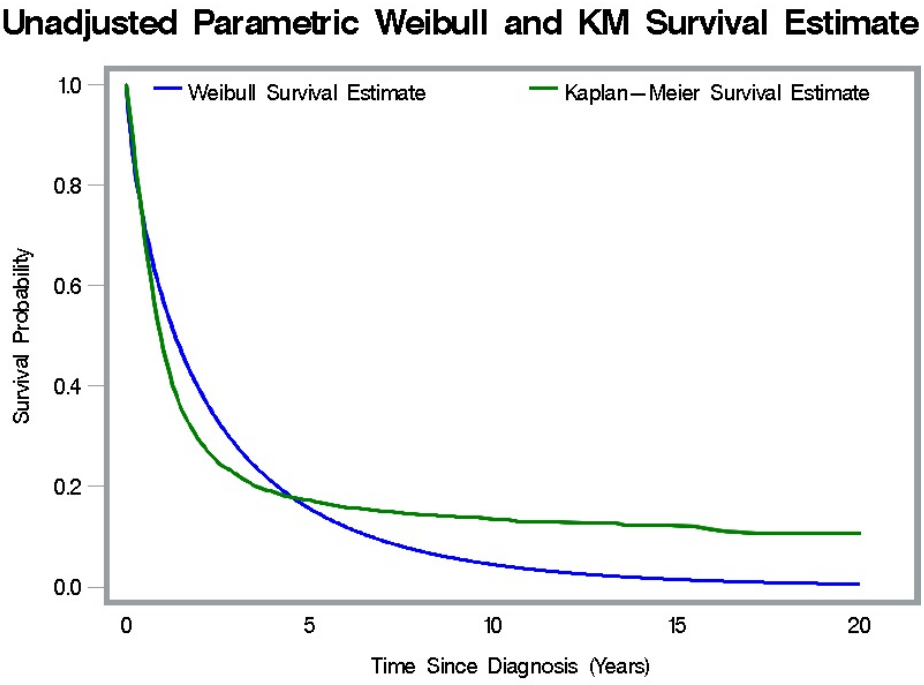

Figure 3.1: The unadjusted survival estimate from the parametric Weibull distribution and the empirical Kaplan-Meier survival function for the SEER esophageal sample.

Tumor length. The influence of tumor length on the survival of esophageal squamous cell carcinoma patients who had surgical resection as the primary treatment has been studied (e.g., Wang et al., 2011b). In this sample of patients, the tumor length variable was defined as the length of the primary tumor (length of involved esophagus ranging from 1 cm to 20 cm). Mean tumor length for patients in the study cohort was 5.2 cm (standard deviation = 2.80, median = 5.0 cm).

The maximum likelihood estimates of the parameters $(\alpha, \lambda, \beta_L)$ were 0.763, 0.340, and 0.095, respectively, as given in Table 3.1 along with the estimated covariance matrix. The LIFEREG procedure in SAS was used to fit a parametric Weibull model with a single continuous covariate. The Wald test for the null hypothesis $H_0 : \beta_L = 0$ versus $H_1 : \beta_L \neq 0$ found that tumor length was a significant predictor of disease-specific survival in the model ($p < .0001$).

Fixing future survival time. Using the maximum likelihood estimates from the survival model, time-conditional survival was estimated using SAS/IML. The time-conditional survival estimator can be written as a function of time survived from diagnosis, a, and future survival time, which is given in Equation 3.11 as

$$CS(b \mid a) = \frac{\exp\left(-b^{\hat{\alpha}} \cdot \hat{\lambda} \cdot \exp\left(\hat{\beta}_L * L\right)\right)}{\exp\left(-a^{\hat{\alpha}} \cdot \hat{\lambda} \cdot \exp\left(\hat{\beta}_L * L\right)\right)},$$

where L is the continuous covariate representing tumor length. Section 3.2.1, and Section B.1 in the Appendix, derive the variance of a time-conditional survival estimator and derive the covariance between any two time-conditional survival estimators from a Weibull regression model with covariates.

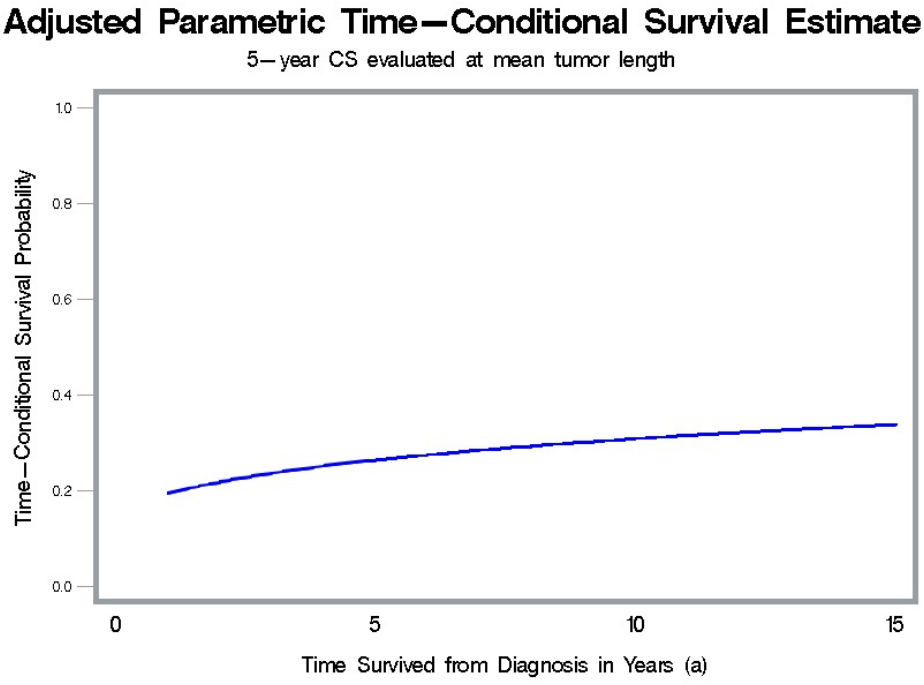

Figure 3.2: Estimated 5-year time-conditional survival probability given increasing time survived for mean tumor length from the Weibull distribution based on the SEER esophageal sample.

Fixing Δ, the estimated 5-year time-conditional survival probability is shown in Figure 3.2 as a function of time survived after diagnosis (a) at the mean value for tumor length (L). The estimated probability of surviving an additional 5 years after diagnosis increased from 0.19 to 0.34 as time survived after diagnosis increased from 1 to 15 years.

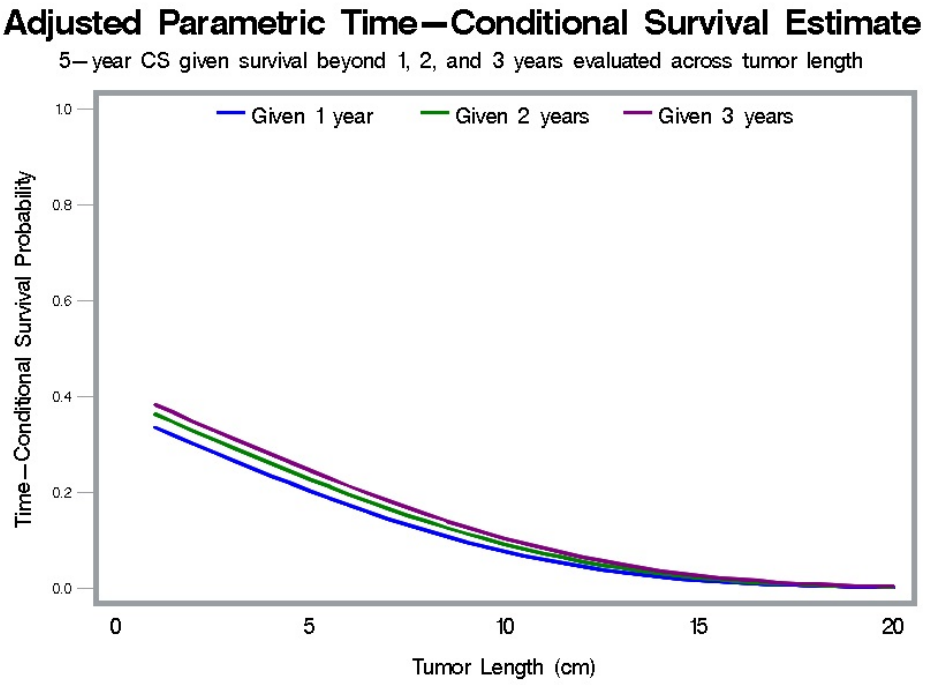

Figure 3.3: Estimated 5-year time-conditional survival probability, given that survival is greater than 1, 2, and 3 years after diagnosis, for increasing tumor length from the Weibull distribution based on the SEER esophageal sample.

Allowing tumor length to increase from 0 to 20 cm, Figure 3.3 plots the estimated 5-year time-conditional survival probability, given that survival is greater than 1, 2, and 3 years after diagnosis. Five-year time-conditional survival probability given survival greater than 1 year after diagnosis dropped from 0.34 to 0.001 with increasing tumor length over the range of 0 to 20 cm. Similarly, the probability dropped from 0.36 to 0.002 and from 0.38 to 0.003, given that survival is greater than 2 and 3 years, respectively. Figure 3.3 plots the estimated 5-year time-conditional survival probability, given that survival is greater than 1, 2, and 3 years after diagnosis, as a function of increasing tumor length. The 5-year time-conditional survival probability given that survival is beyond 3 years was greater than survival beyond 2 years across all values of tumor length. Similarly, 5-year time-conditional survival probability given that survival is beyond 2 years was greater than 5-year time-conditional survival probability given that survival is beyond 1 year after diagnosis across all values

of tumor length.

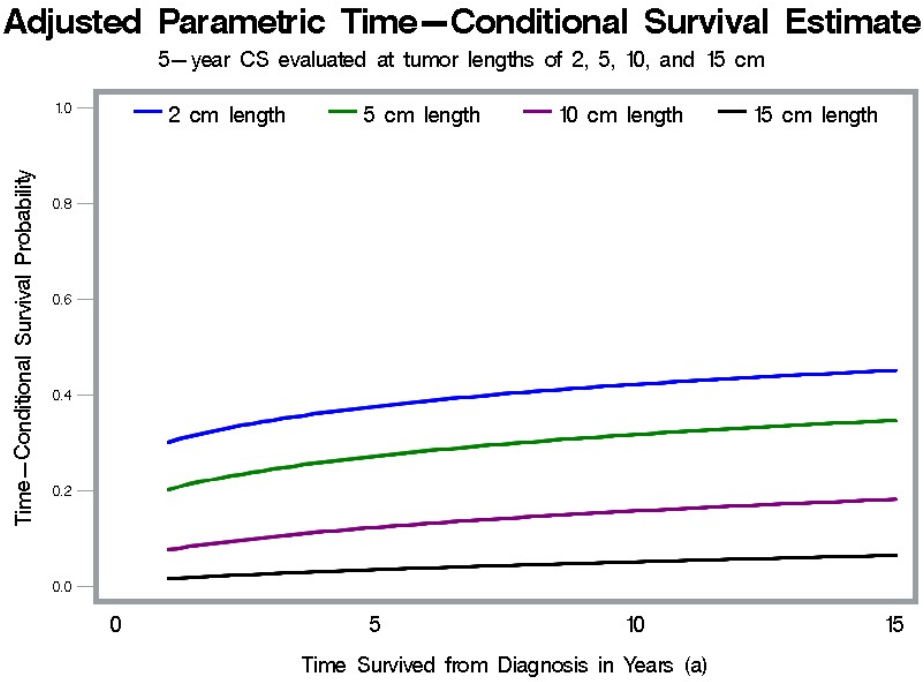

Figure 3.4: Estimated 5-year time-conditional survival probability given increasing time survived for tumor length at 2, 5, 10, and 15 cm from the Weibull distribution based on the SEER esophageal sample.

Including tumor length as a continuous covariate in the estimation of survival probability and in the estimation of time-conditional survival probability allows for the presentation of time-conditional survival estimates by profiles defined by values of the continuous covariate. Figure 3.4 plots the estimated 5-year time-conditional survival probability profile given increasing survival after diagnosis for tumor lengths 2, 5, 10, and 15 cm. At 2 cm in length, the estimated time-conditional survival probability ranged from 0.30 to 0.45 and was consistently greater than the estimates observed for greater tumor lengths (0.20 to 0.35 for 5 cm; 0.08 to 0.18 for 10 cm; and 0.02 to 0.06 for 15 cm). These changes in time-conditional survival probability for tumor lengths 2, 5, 10, and 15 cm are clinically meaningful. For example, a patient who survives 15 years post diagnosis can be told that she will have approximately a 45% likelihood of surviving an additional 5 years, if her tumor length at diagnosis was 2 cm. In contrast, her chances of surviving an additional 5 years will be only approximately 6%, if her tumor length at diagnosis was 15 cm.

Fixing time survived. Figure 3.5 shows the estimated time-conditional survival probability, given

that survival is greater than 2 years after diagnosis, evaluated at mean tumor length. The probability of surviving an additional 1 to 18 years beyond 2 years, given that survival is greater than 2 years, decreased from 0.80 to 0.03. As a function of tumor length, Figure 3.6 plots the estimated 2-, 3-, 4-, and 5-year time-conditional survival probability, given that survival is greater than 2 years after diagnosis, for increasing tumor length.

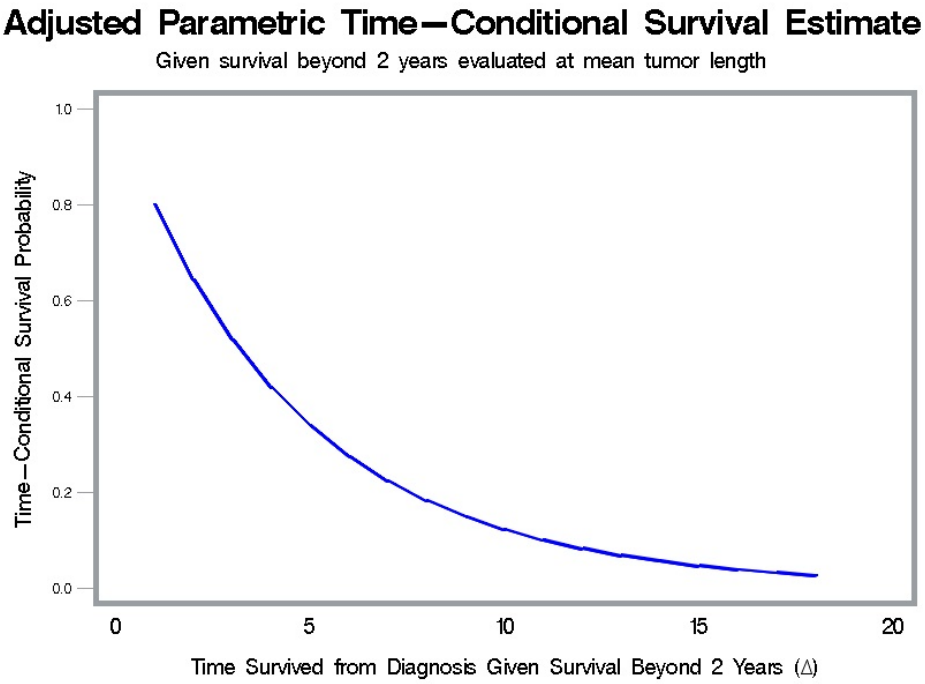

Figure 3.5: Estimated time-conditional survival probability, given that survival is greater than 2 years after diagnosis, as a function of Δ evaluated at mean tumor length from the Weibull distribution based on the SEER esophageal sample.

Given survival 2 years after diagnosis, the probability of surviving an additional 2 years decreased from 0.64 to 0.07, the probability of surviving an additional 3 years decreased from 0.53 to 0.02, the probability of surviving an additional 4 years decreased from 0.44 to 0.01, and the probability of surviving an additional 5 years decreased from 0.36 to 0.002. Across tumor length, 2-year time-conditional survival probability was greater than the 3-year, the 3-year was greater than the 4-year, and the 4-year was greater than the 5-year. Under this parametric framework, the figure shows that the likelihood of surviving additional time beyond 2 years, given that survival is greater than 2 years, decreases irrespective of tumor length. We found that there is no significant difference between 5-year time-conditional survival probability, given that survival is greater than 2 years, as compared to 4-year time-conditional survival probability, given that survival is greater than 2 years

(TS=1.86, p=0.1729).

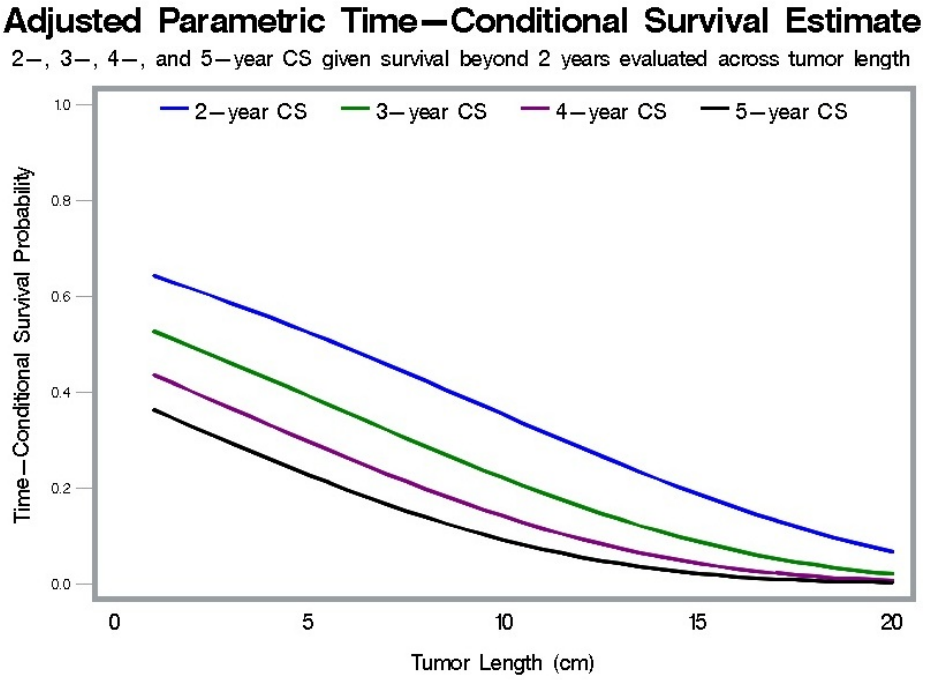

Figure 3.6: Estimated 2-, 3-, 4-, and 5-year time-conditional survival probability, given that survival is greater than 2 years after diagnosis, for increasing tumor length from the Weibull distribution based on the SEER esophageal sample.

Hypothesis testing. To assess whether the likelihood of surviving an additional 5 years beyond 1 year, given that a patient has survived more than 1 year after diagnosis, varies for different tumor lengths, we define CS_1 as the 5-year time-conditional survival for those with a tumor length of 2 cm at diagnosis and CS_2 for those with tumor length of 5 cm at diagnosis. From Equation 3.11, the 5-year time-conditional survival probability given 1 year for those with a tumor length of 2 cm and 5 cm was 0.30 and 0.20, respectively. The estimated covariance matrix is given by

$$\widehat{\boldsymbol{\Sigma}} = \begin{pmatrix} \widehat{Var}(\widehat{CS}_1) = 13.67 & \widehat{Cov}(\widehat{CS}_1, \widehat{CS}_2) = 9.08 \\ \widehat{Cov}(\widehat{CS}_1, \widehat{CS}_2) = 9.08 & \widehat{Var}(\widehat{CS}_2) = 6.75 \end{pmatrix} \times 10^{-5}.$$

These estimates resulted from transforming the Weibull maximum likelihood estimates obtained using the $\mathrm{LIFEREG}$ procedure in SAS.

As shown in Figure 3.4, with increasing tumor length from 2 cm to 15 cm, the time-conditional survival probability decreased. Specifically, with increasing time after diagnosis from 1 year to

15 years, patients with a tumor length of 2 cm had a 0.15 increase in the probability of surviving an additional 5 years (estimated time-conditional survival ranged from 0.30 to 0.45). Alternatively, patients with a tumor length of 15 cm had a 0.04 increase in the probability of surviving an additional 5 years (estimated time-conditional survival ranged from 0.02 at 1 year after diagnosis to 0.06 at 15 years after diagnosis).

We applied the hypothesis test shown in Equation 3.4 to evaluate the influence of tumor length (2 cm vs 5 cm) on 5-year time-conditional survival probabilities, given that survival is greater than 1 year after diagnosis. The χ^2 test statistic calculated from Equation 3.5 was 431.6, indicating a significant difference ($p < 0.0001$). The likelihood of surviving an additional 5 years given survival beyond 1 year after diagnosis is significantly better for those with tumor length of 2 cm as compared with 5 cm. Refer to the Section B.1 of the Appendix for details on the calculations. In conclusion, tumor length at diagnosis is an important continuous covariate to adjust for when estimating time-conditional survival probabilities for patients with esophageal cancer.

3.4.2. Cure Model Based Time-Conditional Survival Estimation for Melanoma

Background. Unlike esophageal cancer, most cases of deaths due to melanoma of the skin are preventable because of its readily recognizable lesions and improved morbidity and mortality due to early detection and intervention (MacKie et al., 1997). As a result, time to disease-specific death in this mixed cohort of individuals who are and are not at risk of dying from melanoma can be studied with a cure model.

Sample. Population-based data from the SEER program were evaluated from a cohort of 3478 melanoma patients diagnosed between 1988 and 2003 and located in the original 9 SEER regions (San Francisco-Oakland, Connecticut, Metropolitan Detroit, Hawaii, Iowa, New Mexico, Seattle (Puget Sound), Utah, and Metropolitan Atlanta). Patients in this cohort had a first (and no other) malignant primary skin melanoma confirmed microscopically (positive histology), tumor thickness ranging from 1.01 to 9.90 mm, and follow-up time greater than zero months (to exclude patients who were diagnosed at autopsy). They all had known age at diagnosis, ulceration status, and number of regional lymph nodes examined. They also had no clinical/pathological lymph node involvement and either had one or more (nonpalpable) nodes examined and determined to be negative or none of their (nonpalpable) nodes examined. Based on the application of these inclusion/exclusion cri-

teria for patients with no clinical/pathological lymph node involvement, this sample represents a well-defined cohort of Stage II patients based on AJCC 6 with long-term follow-up.

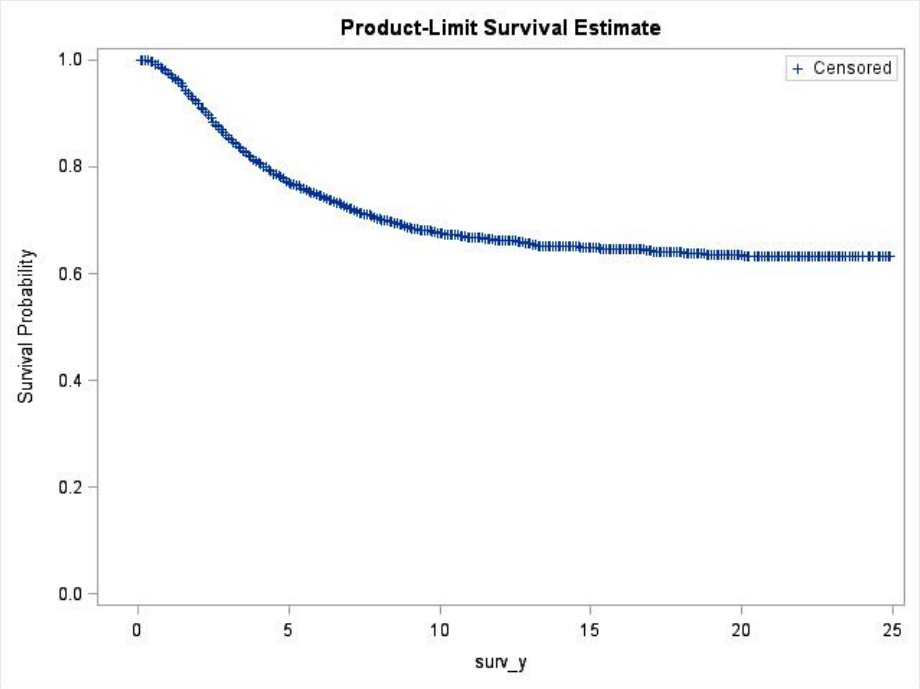

Figure 3.7: Empirical survivor function for the SEER melanoma sample.

Figure 3.7 shows the empirical survival function where the survival probability in year 1 through year 25 after diagnosis is above 0.60. This suggested that a cure model approach would be appropriate for this data analysis. The empirical survival estimate at 10, 15, and 20 years after initial diagnosis was 67.6%, 64.8%, and 63.6%, respectively. This indicated that there were only small changes in the survival curve with respect to disease-specific death for patients who survived 10 years after diagnosis. Typical survival models for these data would assume that the underlying population is susceptible to death due to disease if followed for a long enough period of time. Inspection of the right tail of the figure suggested that this was not the case, which demonstrated the appropriateness of exploring the mixture cure model approach for these data.

The research objective of this analysis was to estimate time-conditional survival probabilities using a Logistic-Weibull cure model to compare the disease-specific survival of patients who are clinical stage II (no regional nodes examined) with and without ulceration to patients who are pathological stage II (nodes examined without evidence of metastasis) with and without ulceration. A measure of

the generalizability of the methodology and hypothesis testing framework described in Section 3.2.3 is demonstrated here by comparing 5-year time-conditional survival probabilities among these four cases of patients with varying staging intensity (clinical or pathological), number of nodes examined, and ulceration status (with or without ulceration) given that survival is greater than 1 and 10 years after diagnosis.

Covariate selection. The selection of appropriate covariates can help to obtain unbiased and accurate estimates of the predictions of time-conditional survival probabilities. The approach to selecting from among those available in SEER the appropriate covariates was based on prior knowledge of the clinical literature. The underlying model framework was made up of two components. The first component modeled the effect of ulceration, stage, gender, and the number of nodes examined on the probability of patients being not cured of melanoma. For the AJCC staging system, the presence or absence of ulceration of the primary tumor is an important prognostic factor in patients with stage I and II disease. Due to the pattern of upstaging by the presence of ulceration in the AJCC 6 staging system, we consider ulceration status as informing the estimation of cure probability. "Staging intensity" was defined by clinical stage II for patients with no regional nodes examined and pathological stage II for patients with 1 or more nodes examined and without evidence of metastasis. As number of positive nodes is necessarily 0 for subjects to fall within this sample, we consider number of nodes examined instead. Albeit indirect, this was used as another measure of aggressiveness of disease in this analysis. As done in Gimotty et al., 2005, we assumed that those patients with no nodes examined had no clinical evidence by palpation of nodal involvement at diagnosis. Therefore, staging intensity and number of nodes examined were important prognostic factors in determining the probability of patients being not cured of melanoma. Lastly, we adjusted for gender as it is available in SEER and has been shown to be an important prognostic factor (for example, Balch, 1992; deVries et al., 2008).

The second component modeled the effect of age at diagnosis and tumor thickness on the survival distribution for patients being not cured of melanoma. A well-known predictor of melanoma outcome is tumor thickness, which is defined as the vertical thickness of the lesion as measured in mm by light microscopy of the biopsy specimen. Tumor thickness is a strong, well validated marker of disease and greater tumor thickness is associated with worse prognosis and survival (Dennis, 1999). Gimotty et al., 2005 note that thickness is the best univariate predictor of disease-

specific survival among patients with melanoma. In their validation of the AJCC melanoma staging, citebalch2001final found that age was statistically significant and an independent measurement associated with clinical outcomes for patients with no evidence of metastatic disease. Therefore, in the estimation of survival time, we include two continuous covariates that are available in SEER, tumor thickness and age at diagnosis, as important factors with impact on survival.

Let the time-conditional survival probability be a function of the following dichotomous and continuous covariates described above: ulceration status (z_1), staging intensity (z_2), gender (z_3), number of nodes examined (z_4), age at diagnosis (x_1), and tumor thickness (x_2). This analysis evaluated the effect of the first set of covariates on the proportion of individuals not cured and the effect of age at diagnosis and tumor thickness on the survival distribution of those not cured. Then estimates of time-conditional, disease-specific survival for stage II patients with varying values of nodes examined, clinical/pathologic staging, ulceration status, gender, age at diagnosis, and tumor thickness were obtained.

Survival estimation. The `%PSPMCM` macro was used to obtain maximum likelihood estimates for the mixture cure model in SAS (Corbière and Joly, 2007). In this macro, the maximization of the likelihood function for the parametric model was carried out using the `PROC NLMIXED` procedure in SAS. The data was entered with one record per patient, which contained observed time (either censoring or failure time), a censoring indicator, and a vector of covariates.

The survivor function for the Logistic-Weibull cure model used for this analysis in Equations 3.3 and 3.12 is given by

$$S(t \mid x_1, x_2, z_1, z_2, z_3, z_4) = \frac{\exp(\eta_0 + \eta_1 z_1 + \eta_2 z_2 + \eta_3 z_3 + \eta_4 z_4)\exp(-t^{\alpha} \cdot \lambda \cdot \exp(\zeta_1 x_1 + \zeta_2 x_2)) + 1}{\exp(\eta_0 + \eta_1 z_1 + \eta_2 z_2 + \eta_3 z_3 + \eta_4 z_4)}.$$

All of the regression coefficients for the covariates were significant in the cure model (see Table 3.2). Based on this model, the likelihood that a patient would not be cured was higher if their tumor had ulceration as compared to patients without ulceration ($\hat{\eta}_1 = 0.335, p = 0.0001$), they were clinical stage II as compared to pathological stage II ($\hat{\eta}_2 = 0.636, p < 0.0001$), and they were males as compared to females ($\hat{\eta}_3 = 0.503, p < 0.0001$). The likelihood of not being cured increased with increasing number of nodes examined ($\hat{\eta}_4 = 0.014, p = 0.0314$).

Similarly, both tumor thickness and age at diagnosis influenced the survival of the not cured patients. There was significant evidence that disease-specific survival among patients who were not cured was worse with increasing age at diagnosis ($\hat{\zeta}_1 = 0.009, p = 0.0012$) and with increasing tumor thickness ($\hat{\zeta}_2 = 0.092, p < 0.0001$). The estimated covariance matrix for the parameters in this model is presented in Table 3.3. This estimated matrix is important for hypothesis testing and evaluation of time-conditional survival probabilities estimated using this model because it is involved in the computation of the test statistic.

Cure proportion. In general terminology, the cure proportion is given by $1 - \hat{\pi}(\mathbf{Z})$ where

$$\hat{\pi}(\mathbf{Z} = \mathbf{z}) = \frac{\exp(\hat{\boldsymbol{\beta}}\mathbf{z})}{1 + \exp(\hat{\boldsymbol{\beta}}\mathbf{z})},$$

is the estimated probability of the event of being not cured for those with $\mathbf{Z} = \mathbf{z}$. The proportion of cured patients is a function of ulceration status, stage, gender, and the number of nodes examined. For this sample of 3478 patients, the estimated cure proportion ranges from approximately 0.41 to 0.79 (mean = 0.64; median = 0.66). For example, the cure proportion for males with clinical stage II disease (no nodes examined) and an ulcerated lesion was 0.46 and was smaller than the 0.61 cure proportion for males with pathological stage II, 1 node examined, and an ulcerated lesion.

CS estimation. From the output of the %PSPMCM macro used to obtain maximum likelihood estimates for the mixture cure model in SAS (Corbière and Joly, 2007), maximum likelihood estimates were transformed to the desired Weibull parametrization. The parameter transformations are given by $\alpha = \frac{1}{\sigma}, \lambda = \exp(-\frac{\mu}{\sigma})$, and $\zeta_i = -\frac{\gamma_i}{\sigma}$ from the Weibull model. Under this transformation, as in Equation 3.13, the time-conditional survival is defined by

$$\begin{aligned} &CS(a + \Delta \mid a, x_1, x_2, z_1, z_2, z_3, z_4) \\ &\qquad = \frac{\exp(\eta_0 + \eta_1 z_1 + \eta_2 z_2 + \eta_3 z_3 + \eta_4 z_4) \exp(-(a + \Delta)^\alpha \cdot \lambda \cdot \exp(\zeta_1 x_1 + \zeta_2 x_2)) + 1}{\exp(\eta_0 + \eta_1 z_1 + \eta_2 z_2 + \eta_3 z_3 + \eta_4 z_4) \exp(-a^\alpha \cdot \lambda \cdot \exp(\zeta_1 x_1 + \zeta_2 x_2)) + 1}. \end{aligned} \tag{3.14}$$

The variance of a time-conditional survival probability estimated from a cure model of a similar form and the partial derivatives are found in Sections B.1 and B.2 of the Appendix.

Hypothesis testing. When a patient is evaluated, the clinician/researcher is interested in how the patient and disease characteristics relate to survival. To evaluate this, she would estimate the survival probability for that patient. Both patient and physician are also interested in understanding how the probability of survival for an additional 5-years (Δ) changes the more time passes after

diagnosis (further increasing a). This requires the estimation of time-conditional survival probabilities based on the information and model relevant to that patient. In general, upon obtaining an appropriate model for survival estimation and estimating the time-conditional survival, the hypothesis testing framework is used to compare point estimates that account for the correlation in the estimation of time-conditional survival probability. The question of whether profiles based on these continuous covariates are significantly different was evaluated by fixing some of the covariates and varying others to produce four time-conditional survival probability profiles.

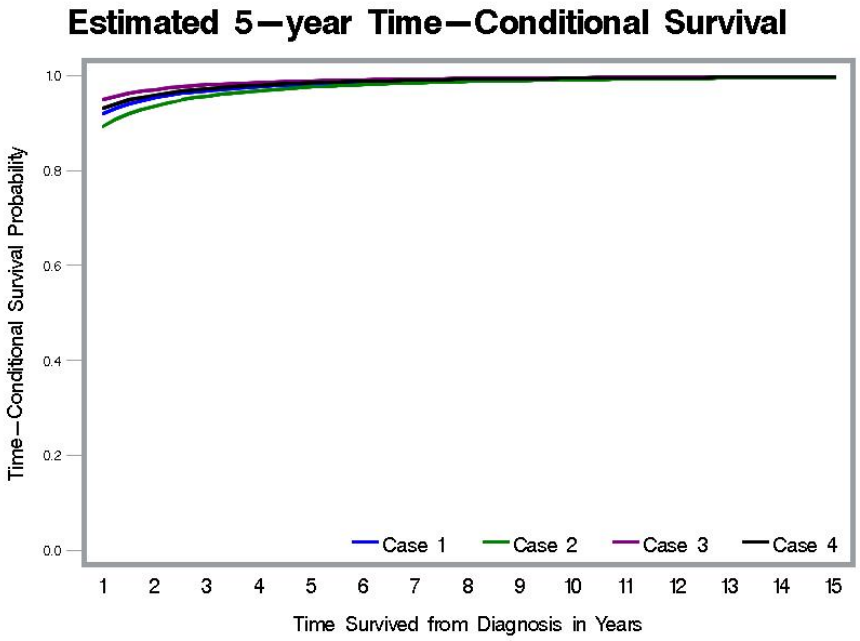

Figure 3.8: Estimated 5-year time-conditional survival for the SEER melanoma sample with four cases based on the logistic-Weibull cure model.

To address the research question around the potentially different survival of those with no palpable nodes (clinical staging), with and without tumor ulceration, versus histologically proven negative nodes (pathological staging), with and without ulceration, we evaluated 5-year time-conditional survival given survival beyond 1 year versus beyond 10 years after diagnosis for disease-specific survival based on several patient and disease characteristics. Four cases were defined to represent patients with increasing disease severity hypothesized to result in worse survival. Each case had varying values for the continuous and dichotomous covariates incorporated into the model to demonstrate the statistical generalizability and usefulness of the approach.

Table 3.4 lays out the characteristics of these four cases defined as the following. Three of the

model covariates were fixed: gender was fixed to be male, age at diagnosis was set to the mean of the sample (60 years old), and tumor thickness was fixed to the mean of the sample at 3.58 mm. The number of nodes examined was set to 0 (clinical stage II) or 1 (pathological stage II), where cases with 1 node examined may be patients who have had a sentinel node biopsy to examine the first node to be involved in lymphatic spread. Case 1 was representative of patients without ulcerated tumors and clinical stage II (no nodes examined). Case 2 was representative of patients with ulceration and clinical stage II (no nodes examined). Case 3 was representative of patients without ulceration, pathological stage II disease and 1 node examined. Lastly, Case 4 was representative of patients with ulceration, pathological stage II disease and 1 node examined. We would expect the 5-year time-conditional survival probabilities given 1 and given 10 years after diagnosis to be different upon evaluating these cases.

Figure 3.8 shows 5-year time-conditional survival probability estimates for up to 15 years after diagnosis across the four cases. To address the question of whether at least one of the time-conditional survival estimates for these four cases was different at 1 year after diagnosis as compared to at 10 years after diagnosis, we used a model similar to that given in Equation 3.14. Define CS_{ij} as the 5-year time-conditional survival probability for Case i at $a = 1$ ($j = 1$) and $a = 10$ ($j = 2$) years after diagnosis ($i = 1, \ldots, 4$). Then, the vector of time-conditional survival probability estimates is given by

$$(CS_{11}, CS_{12}, CS_{21}, CS_{22}, CS_{31}, CS_{32}, CS_{41}, CS_{42})^T,$$

and the estimates are given in Table 3.4.

A multivariate test of pairwise differences was used for this analysis to address the research question and to avoid the problem of multiple testing when evaluating each pairwise comparison individually. The null hypothesis is that the 5-year time-conditional survival is the same given survival beyond 1 year versus beyond 10 years after diagnosis in each of the four cases representing patients with no palpable nodes (clinical staging), with and without ulceration, and negative nodes

(pathological staging), with and without ulceration. This is given by

$$H_0 : \begin{array}{l} CS_{11} - CS_{12} = 0 \\ CS_{21} - CS_{22} = 0 \\ CS_{31} - CS_{32} = 0 \\ CS_{41} - CS_{42} = 0 \end{array}$$

and can be written in the form $\mathbf{XCS}$ as above. In this hypothesis testing setting, we adjusted for the correlation among the differences. Here, the test statistic has the same functional form as in Equation 3.8. The original covariance matrix is shown in Table 3.3 and the estimated covariance matrix for the differences shown in Table 3.4 has elements based on the computations from Appendix B.2. Additional discussion on model building, covariate selection, and the impact on estimated time-conditional survival probabilities is in Appendix B.3.

From Table 3.4, the estimated differences in 5-year time-conditional survival probabilities given survival beyond 1 versus beyond 10 years were 0.073 for Case 1, 0.098 for Case 2, 0.041 for Case 3 and 0.056 for Case 4. All of these estimates were positive and this indicated that, across all four cases, estimated 5-year time-conditional survival probabilities, given that survival was beyond 10 years after diagnosis was greater than given 1 year after diagnosis. For example, the range in estimated differences indicated that those patients with pathological stage II disease without ulceration had better 5-year prognosis both initially and 10 years after diagnosis as compared to those with clinical stage II disease and ulceration (Case 3: $\widehat{CS}(6 \mid 1) = 0.955$ and $\widehat{CS}(15 \mid 10) = 0.997$; Case 2: $\widehat{CS}(6 \mid 1) = 0.893$ and $\widehat{CS}(15 \mid 10) = 0.991$). Assessing the long-term time-conditional survival, we found that pathological stage II patients have improved long-term disease-specific survival as compared to similar clinical stage II patients, irrespective of ulceration status with 95% confidence intervals that account for the correlation of the time-conditional survival estimates for Case 1 (0.0285, 0.1175), for Case 2 (0.0435, 0.1525), for Case 3 (0.0001, 0.0819), and for Case 4 (0.0125, 0.0995), respectively.

Based on the hypothesis test defined above, the test statistic was estimated to be 33.4 with degrees of freedom equal to $Rank(X) = 4$. There was significant evidence to indicate that at least one of the 5-year time-conditional survival probability estimates, given that survival is greater than 1 year as compared to 10 years after diagnosis, was different among Cases 1, 2, 3, and 4 ($p <$

0.0001). Further, we investigated the independent pairwise contrasts. Table 3.5 shows the pairwise comparisons for the change in 5-year time-conditional survival probability given 1 and given 10 years after diagnosis for each case with Bonferroni adjustment. The unadjusted p-values show that all four cases are significant at the 0.05 α-level. Upon adjusting for multiple comparisons, three of the four tests were significant. There was statistically significant change in 5-year time-conditional survival probability between 1 year after diagnosis and 10 years after diagnosis for the cases representing clinical stage II without ulceration ($p = 0.0052$), clinical stage II with ulceration ($p = 0.0020$), and pathological stage II with ulceration ($p = 0.0448$). The estimated change in 5-year time-conditional survival was not significant for the case representing pathological stage II without ulceration (Bonferroni-adjusted p-value = 0.1916).

Overall, our analysis of the four cases found that at least one of the cases of pathological stage II versus clinical stage II patients by ulceration status had improved long-term disease-specific survival ($p < 0.0001$). From the pairwise contrasts, we found a significant increase in 5-year time-conditional survival probability estimates for both cases of clinical stage II (with and without ulceration) and for pathological stage II with ulceration, after the Bonferroni adjustment. The case for pathological stage II without ulceration had the largest 5-year time-conditional survival probability estimates given both 1 and 10 years after diagnosis compared to the other cases and the incremental estimated change was not significant.

3.5. Discussion

This chapter developed and demonstrated a methodology for the inclusion of multiple covariates, including continuous covariates, in the estimation and analysis of parametric time-conditional survival probability. The approach for including continuous covariates was described under the Weibull regression model and the Logistic-Weibull cure model. The Weibull regression model was applied to SEER esophageal cancer data and the cure model was applied to stage-specific SEER melanoma data. We note that if a covariate is significant in the parametric survival model, it remains significant in the time-conditional survival probability model reflecting the direct one to one relationship between survival probability and time-conditional survival probability. While this parametric methodology is applied to SEER data, it could also be applied to any data with information on the long-term follow-up of time to an event. Deriving the approximate large sample distribu-

tion under each of these models allows the researcher to build time-conditional survival probability profiles and use Wald χ^2 test statistics to evaluate differences in estimates.

Using the data on patients with esophageal cancer, we fit a parametric Weibull regression model assessing time to death due to this malignancy as a function of tumor length. There was significant evidence to indicate that tumor length was a predictor of disease-specific survival in the survival model, such that those with shorter tumor lengths had a better probability of survival ($p < .0001$). Using the maximum likelihood estimates from the parametric survival model, time-conditional survival probability was expressed as a function of continuous tumor length, Weibull parameter estimates, time survived, and future survival time. Including tumor length as a continuous covariate in the estimation of parametric survival and subsequently in the estimation of time-conditional survival probability allowed for the consideration of profiles defined by values of the continuous covariate. Tumor length-specific profiles captured the relationship between future survival time and 5-year time-conditional survival probability for any possible values of the continuous covariate. We concluded that tumor length at diagnosis was an important continuous covariate to adjust for when estimating time-conditional survival probabilities for patients with esophageal cancer.

For the melanoma data, we used the hypothesis testing framework to compare point estimates that accounted for the correlation in the estimation of time-conditional survival probability from a mixture cure model. Some of the questions evaluated in this real-world application include: (1) understanding how the probability that a patient survived for an additional 5 years (Δ) changed the more time passed after diagnosis (further increasing a), (2) how patient age and tumor thickness at diagnosis related to survival, and (3) whether estimates for the four cases described were different 1 year after diagnosis as compared with 10 years after diagnosis. Adjusting for patient and disease characteristics, we found that a significantly larger proportion of patients were not cured when the number of nodes examined increased. Further, the analysis indicated that disease-specific survival among patients who were not cured worsened with increasing tumor thickness and with increasing age. Lastly, we assessed several cases with respect to 5-year time-conditional survival probability estimates to demonstrate the flexibility of the hypothesis testing framework. We found that patients who were pathological stage II (nodes examined and without evidence of metastasis) had better long-term time-conditional disease-specific survival probability as compared to patients who were clinical stage II (no regional nodes examined). Specifically, given that survival is greater than 1 year

as compared to greater than 10 years after diagnosis, there was significant evidence to indicate that at least one of the estimates of 5-year time conditional survival probability among the four cases was different ($p < 0.0001$).

There are some limitations to using methods that depend on the assumption of a parametric distribution. The use of parametric models may lead to more precise estimates as compared to their nonparametric counterparts due to the loss of information and reduced power of a test that ignores the variability within strata (Greenland, 1995). On the other hand, it is important to note that model misspecification may lead to consistent estimation of a biased estimator. Although the parametric assumptions may simplify a complex underlying disease mechanism, the proposed methodology is highly useful in the circumstances where there is an adequate fit of the model to the data. The cost of this additional information is, of course, the requirement that some parametric assumptions be made. If these assumptions are not valid, then the cure model results are inaccurate. As with other regression approaches, when comparing profiles the researcher must take care to avoid extrapolating above and beyond the sample on which the model was built when looking to evaluate realistic profiles.

Researchers must also be cautious when computing estimates of survival probability based on sparse data. Such situations may result in larger variances and will directly impact the calculation of the test statistic, which will lead to decreased power to reject the null hypothesis. Approaching the problem from the perspective of the cure model provides additional information by comparing groups in terms of both the proportion “cured” and the survival of those not “cured.” Further, looking at time-conditional survival may be of greater interest among the subset of patients with a high enough risk of death under the cure model. Likely the long-term follow-up among those who with a higher probability of being cured will be less informative as their time-conditional survival probability will be almost 1.

Researchers implementing methods in time-conditional survival probability should take care to ensure inverse stability in their computations. Our code incorporated several checks to assess the potential inverse stability and that there were no computational problems such as mutli-collinearity. We included two types of checks to ensure that the covariance matrix used was stable. First, a user-defined function was used to compute the rank of the covariance matrix. The function was written to be able to compute the rank of any matrix by performing elementary row operations to

get the matrix into echelon form using the built-in SAS/IML command $\mathrm{echelon}()$ (SAS Institute Inc., 2008). Then, in echelon form, the number of non-zero rows is the rank of the matrix. The rank is computed for the vector of time-conditional survival probabilities, the design matrix for the relevant hypothesis test, and the covariance matrix of the differences. The second check occurred during the estimation of the inverse of the covariance matrix. The inverse of the covariance matrix is necessary to compute the test statistic for the hypothesis tests defined in this chapter. The built-in function SAS/IML function $\mathrm{inv}()$ is used to compute the inverse (SAS Institute Inc., 2008). This function computes the inverse of a square, non-singular matrix. The matrix must be square and non-singular for this function to run. If the matrix is singular, then no output will be produced in SAS, an error will be produced in the log file, and the test statistic will not be computed.

As with the nonparametric methods, parametric time-conditional survival probability estimates can help with shared decision making by patient and physician by providing relevant information on future survival. In moving towards personalized medicine, the parametric estimation approach evaluates time-conditional survival as a function of both categorical and continuous patient and disease characteristics at continuous times after diagnosis. Survival statistics are of interest when providing estimates of prognosis but most data are based on projections from survival at diagnosis of a specific cohort. As noted by Bryant et al., 2012, a patient's journey may improve with cancer care and their survival probability may change and an example of this was shown in a wide range of cancer diagnoses among Canadian patients (Ellison et al., 2011). This relationship can be assessed for any possible value of future survival, given survival is greater than the current survival time.

Time-conditional survival probability addresses this by providing estimates of the probability of surviving beyond a given time point in the future having survived beyond various milestones of the cancer experience. Estimates from parametric time-conditional survival allow the clinician/investigator to adjust for continuous and discrete covariates in the calculation of time-conditional survival relative to a specific set of patient characteristics of interest. Specifically, these estimates can be personalized by allowing customization of the estimate of interest with the exact values relevant to an individual patient.

Table 3.1: Maximum likelihood estimates of the Weibull survival distribution for disease-specific survival of esophageal cancer patients adjusting for tumor length.

Parameters	Estimates	Covariance Matrix		
α	0.763	4.84×10^{-5}	-2.08×10^{-5}	3.37×10^{-6}
λ	0.340	-2.08×10^{-5}	1.65×10^{-5}	-4.06×10^{-5}
β_L	0.095	3.37×10^{-6}	-4.06×10^{-5}	1.53×10^{-5}

Table 3.2: Maximum likelihood estimates of the Weibull mixture cure model for disease-specific survival.

Variable	Estimate	Standard Error	p-value
	Logistic Model		
Intercept	-1.320	0.0934	< 0.0001
Ulceration (vs without)	0.335	0.0879	0.0001
Clinical Staging (vs Pathological)	0.636	0.0929	< 0.0001
Males (vs Females)	0.503	0.0832	< 0.0001
No. Nodes Examined	0.014	0.0063	0.0314
	Weibull Survival Model		
Intercept (Weibull)	2.262	0.1216	< 0.0001
Age at Diagnosis (years)	0.009	0.0026	0.0012
Thickness (mm)	0.092	0.0188	< 0.0001
Shape (Weibull)	0.684	0.0178	< 0.0001

Table 3.3: Estimated covariance matrix* for the time-conditional survival probability from the Weibull mixture cure model for disease-specific survival.

Intercept	Ulceration	Staging	Gender	No. Nodes Examined	Intercept (Weibull)	Age (Years)	Tumor (mm) Thickness	Shape (Weibull)
8.714	-2.620	-5.360	-4.250	-0.280	0.930	0.008	0.092	0.100
-2.620	7.729	0.713	-0.180	0.034	0.240	-0.004	0.090	0.037
-5.360	0.713	8.637	0.196	0.283	-0.990	-0.020	-0.050	-0.050
-4.250	-0.180	0.196	6.925	-0.020	-0.070	-0.003	0.002	0.010
-0.280	0.034	0.283	-0.020	0.040	-0.020	-0.025†	-0.002	-0.002
0.930	0.240	-0.990	-0.070	-0.020	14.790	0.274	0.878	0.149
0.008	-0.004	-0.020	-0.003	-0.025†	0.274	0.007	-0.003	-0.005
0.092	0.090	-0.050	0.002	-0.002	0.878	-0.003	0.355	0.012
0.100	0.037	-0.050	0.010	-0.002	0.149	-0.005	0.012	0.315

* Original variances and covariances were multiplied by a factor of 10^3 † Multiplied by a factor of 10^5

Table 3.4: Estimates of 5-year time-conditional survival probability from the Weibull mixture cure model for disease-specific survival adjusting for fixed gender (male), fixed age at diagnosis (60 years), fixed tumor thickness (3.58 mm) and varying staging type (clinical versus pathological), number of nodes examined, and ulceration status along with the estimated covariance and correlation matrices for the alternative hypothesis.

Case i	Staging Type	No. Nodes Examined	Ulceration Status	$\widehat{CS}(6 \mid 1)$ $(j = 1)$	$\widehat{CS}(15 \mid 10)$ $(j = 2)$	Estimates (H_1)
1	Clinical	0	Without Ulceration	0.921	0.994	0.073
2	Clinical	0	With Ulceration	0.893	0.991	0.098
3	Pathologic	1	Without Ulceration	0.955	0.997	0.041
4	Pathologic	1	With Ulceration	0.939	0.995	0.056

Case i	Estimated Matrix*			
1	5.150	5.929	3.475	3.904
2	0.939	7.736	3.602	5.325
3	0.734	0.621	4.354	3.730
4	0.775	0.863	0.806	4.924

* Covariances and variances of differences of time-conditional survival estimates are presented in the upper off-diagonal and diagonal elements, respectively, and original variances and covariances were multiplied by a factor of 10^4. Lower off-diagonal elements are correlations among estimates.

Table 3.5: Change in 5-year time-conditional survival probability given 1 and given 10 years after diagnosis from the Weibull mixture cure model for disease-specific survival with Bonferroni adjustment.

Case i	Absolute Change	Estimated Variance*	Test Statistic	Unadjusted p-value	Bonferroni-adjusted p-value
1	0.073	5.150	10.33	0.0013	0.0052
2	0.098	7.736	12.31	0.0005	0.0020
3	0.041	4.354	3.92	0.0479	0.1916
4	0.056	4.924	6.44	0.0112	0.0448

*Variance estimates were multiplied by a factor of 10^4.

CHAPTER 4

ANALYSIS OF LONGITUDINAL COUNT DATA WITH SPECIFIED MARGINAL MEANS AND FIRST-ORDER ANTEDEPENDENCE

4.1. Introduction

Longitudinal count data are often encountered in scientific studies. For example, Thall and Vail, 1990 analyzed repeated seizure counts on subjects in a clinical trial. Common features of serial count data include intra-subject correlation that is due to similarity between the repeated measurements on each participant. Over-dispersion, which occurs when the variance is larger than expected for the assumed distribution of the outcome variable, is another common feature of longitudinal count data (Efron, 1992). Poisson regression is often applied for analysis of count data but is usually not appropriate for longitudinal studies because it ignores intra-subject correlations and over-dispersion. Generalized Poisson regression (Consul and Famoye, 1992) allows for both over and under dispersion but assumes independence of measurements.

In this chapter we implement a maximum-likelihood based method for analysis of longitudinal count data with over-dispersion that is induced by the serial correlation of measurements. Key assumptions of the approach include the first-order Markov property and linearity of the expectations for the conditional distributions, which are assumed to be Poisson. In addition, we assume that the correlation between adjacent measurements on a subject is constant.

The assumptions of the first-order Markov property, linearity in the conditional expectations, and constant adjacent correlations have been shown to induce a first-order autoregressive AR(1) correlation structure for the repeated outcomes on each subject (Guerra and Shults, 2014). The AR(1) structure is often applied for analysis of data that are equally spaced in time because the assumption of constant adjacent correlations is most plausible when the temporal spacing of consecutive measurements is constant. The AR(1) structure also forces a decline in the intra-subject correlations with increasing separation in time that is plausible for longitudinal trials. Our method is therefore most appropriate for analysis of equally spaced longitudinal count data with over-dispersion.

Other approaches for analysis of over-dispersed longitudinal count data include semi-parametric

approaches such as generalized estimating equations (GEE) (Liang and Zeger, 1986). GEE is widely used because it does not require specification of the full likelihood that can be quite complex for longitudinal discrete data. However, GEE does not account for over-dispersion. In addition, the relative ease of application for GEE can also be a potential limitation for the approach for discrete data. When only the first two moments of the distribution of the outcome variable are estimated, as they are for GEE, it is possible to obtain estimates that are not compatible with any valid parent distribution. In other words, it is possible to obtain estimates for which no corresponding outcome distribution can be constructed. As cautioned by Molenberghs and Kenward, 2010, "the parent provides a natural description of the framework into which the semi-parametrically specified parameters fit. The implication is that such semi-parametric methods as GEE1, GEE2, ALR, etc. can always be applied because there is always a valid parent, and hence a probabilistic basis."

We will make comparisons with GEE because GEE is widely used for analysis of longitudinal discrete data. We will also use GEE to obtain *starting values* for estimation. However, we will confirm that the GEE estimates of the correlation parameters satisfy constraints that are compatible with a valid parent distribution. We conduct simulations for moderately sized samples to demonstrate that when the likelihood is correctly specified, we have improved efficiency in estimation of the regression and correlation parameters for our approach relative to GEE.

Other models for longitudinal count data include generalized linear mixed-effects models that incorporate random effects in the linear predictor. However, the implementation of likelihood based methods that involve random effects can be computationally challenging (p. 75 Fitzmaurice et al., 2008). In addition and in contrast to GEE, for mixed models it is not straightforward to specify a particular working correlation structure for the repeated measurements on subjects. For example, the AR(1) correlation structure is not among the covariance models that were suggested by Thall and Vail, 1990. Mixed-effects models are typically employed when the goal is to estimate effects that are subject specific, because the analysis results are conditional on the random effects (Gardiner, Luo, and Roman, 2009).

In general, likelihood based approaches like the one we implement in this chapter enjoy several general advantages. Unlike semi-parametric approaches, they yield an estimated likelihood that can be used to conduct likelihood ratio tests and to compare the fit of nested models using criteria such as the Akaike information (AIC) (Akaike, 1974) and Bayesian information (BIC) (Schwarz,

1978) criteria. Maximum likelihood estimators are also most (asymptotically) efficient among a wide class of estimators (Serfling, 2011). Our method in particular allows for specification of the usual model for the marginal mean for Poisson data, while also accounting for over-dispersion and serial correlation in the data via an induced AR(1) correlation structure.

In Section 4.2 we discuss the notation, model assumptions, the likelihood and likelihood equations. We discuss an application of the methods in Section 4.3 followed by the simulation studies in Section 4.4. We conclude with a discussion in Section 4.5.

4.2. Methods

4.2.1. Notation and Model Assumptions

The data comprise realizations y_{ij} of ordered discrete random variables Y_{ij} that are measured on subject i at time t_{ij} ($i = 1, \ldots, m$ and $j = 1, \ldots, n_i$). Associated with each y_{ij} is a vector of explanatory variables (covariates) $x_{ij} = (x_{ij1}, \ldots, x_{ijp})'$. The expected value of measurement Y_{ij} on subject i is given by

$$E(Y_{ij}) = \mu_{ij} = \lambda_{ij}, \tag{4.1}$$

and the variance by var$(Y_{ij}) = \sigma^2_{ij}$.

We assume that observations on different subjects are independent. Further, the measurements within subjects are correlated with a structure that depends on parameter α. Let cov(Y_{ij}, Y_{ik}) represent the covariance and corr(Y_{ij}, Y_{ik}) represent the correlation between Y_{ij} and Y_{ik}.

We make three assumptions. First, we assume first-order antedependence, such that each Y_{ij}, given the immediate antecedent Y_{ij-1}, is independent of all further preceding variables (Gabriel, 1962). The joint probability mass function of $Y_{i1}, \ldots, Y_{in_i}$ can then be expressed as

$$\begin{aligned} &\mathsf{P}(Y_{i1} = y_{i1}, Y_{i2} = y_{i2}, \ldots, Y_{in_i} = y_{in_i}) = \\ &\quad \mathsf{P}(Y_{i1} = y_{i1})\mathsf{P}(Y_{i2} = y_{i2}|Y_{i1} = y_{i1}) \cdots \mathsf{P}(Y_{in_i} = y_{in_i}|Y_{in-1} = y_{in-1}). \end{aligned} \tag{4.2}$$

First-order antedependence is also referred to as the first-order Markov property in the literature (Feller, 1968, p. 419).

Second, we assume that the correlation between adjacent measurements on a subject is constant, implying that

$$\text{corr}(Y_{ij}, Y_{ij-1}) = \alpha$$

where $i = 1 \ldots, m$ and $j = 2, \ldots, n_i$. Third, we assume that the conditional expectation of Y_{ij} given Y_{ij-1} is a linear function of Y_{ij-1}, such that

$$\text{E}(Y_{ij} \mid Y_{ij-1}) = a_{ij} + b_{ij} Y_{ij-1},$$

for $i = 1 \ldots, m$ and $j = 2, \ldots, n_i$.

These three assumptions imply the following results. From Theorem 2.1 of Guerra and Shults, 2014, the conditional expectation is given by

$$\text{E}\left(Y_{ij} | Y_{ij-1}\right) = \mu_{ij} + \alpha\sigma_{ij}/\sigma_{ij-1}\left(Y_{ij-1} - \mu_{ij-1}\right), \tag{4.3}$$

where $\mu_{ij} = \text{E}(Y_{ij})$, $\alpha = \text{corr}(Y_{ij-1}, Y_{ij})$, ${\sigma_{ij}}^2 = \text{var}(Y_{ij})$, and

$${\sigma_{ij}}^2 = \frac{1}{1-\alpha^2}\text{E}(\text{var}(Y_{ij} \mid Y_{ij-1})), \tag{4.4}$$

where $i = 1, \ldots, m$ and $j = 2, \ldots, n_i$.

Next, from Theorem 2.2 of Guerra and Shults, 2014, the correlation $\text{corr}(Y_{ij}, Y_{ij+t})$ between Y_{ij} and Y_{ij+t} for $t > 0$ can be expressed as

$$\begin{aligned} \text{corr}(Y_{ij}, Y_{ij+t}) &= \prod_{k=j}^{j+t-1} \text{corr}(Y_{ij}, Y_{ij+1}) \\ &= \prod_{k=j}^{j+t-1} \alpha \\ &= \alpha^t. \end{aligned}$$

The induced correlation structure for $(Y_{i1}, \ldots, Y_{in_i})'$ is therefore an AR(1) structure.

This AR(1) structure is plausible for longitudinal data because it requires the correlation between measurements on a subject to decline with increasing separation in time. For example, if $\alpha = 0.5$,

then the correlation between the 1st and 2nd measurements is 0.5, while the correlation between 1st and 3rd measurements is $(0.5)^2 = 0.25$.

4.2.2. Poisson Likelihood

We assume Poisson distributions for the marginal and conditional distributions in Equation 4.2. For each $i = 1, \ldots, m$, the distribution of Y_{i1} is Poisson with $\mu_{i1} = \lambda_{i1} = \exp(x'_{i1}\beta)$ and ${\sigma_{i1}}^2 = \lambda_{i1}$, where β is a $p \times 1$ vector of regression parameters. Then, for $j = 2, \ldots, n_i$, the *conditional* distribution of Y_{ij} given Y_{ij-1} is Poisson with conditional mean $\mathrm{E}(Y_{ij}|Y_{ij-1}) = {\lambda_{ij}}^*$ given by Equation 4.3, with

$$\mu_{ij} = \lambda_{ij} = \exp\left(x'_{ij}\beta\right), \tag{4.5}$$

and

$${\sigma_{ij}}^2 = \lambda_{ij}/(1-\alpha^2), \tag{4.6}$$

for $j = 2, \ldots, n_i$ and $i = 1, \ldots, m$. The Y_{ij} are over-dispersed relative to the Poisson distribution if $j \geq 2$ and $\alpha \neq 0$, because in this case ${\sigma_{ij}}^2 = \phi_{ij}\lambda_{ij}$, where $\phi_{ij} > 1$.

The likelihood can then be expressed as

$$\begin{aligned} L(\beta, \alpha) &= \prod_{i=1}^{m} \mathrm{P}(Y_{i1} = y_{i1})\mathrm{P}(Y_{i2} = y_{i2}|Y_{i1} = y_{i1}) \cdots \mathrm{P}(Y_{in_i} = y_{in_i}|Y_{in-1} = y_{in-1}) \\ &= \prod_{i=1}^{m} \frac{\exp(-\lambda_{i1}){\lambda_{i1}}^{y_{i1}}}{y_{i1}!} \prod_{j=2}^{n_i} \frac{\exp(-{\lambda_{ij}}^*)({\lambda_{ij}}^*)^{y_{ij}}}{y_{ij}!} \\ &= \prod_{i=1}^{m} \exp\left(y_{i1}\ln(\lambda_{i1}) - \lambda_{i1} - \ln(y_{i1}!)\right) \prod_{j=2}^{n_i} \exp\left(y_{ij}\ln({\lambda_{ij}}^*) - {\lambda_{ij}}^* - \ln(y_{ij}!)\right). \end{aligned}$$

Taking the natural logarithm then yields the log-likelihood,

$$\ln\left(L(\beta, \alpha)\right) = \sum_{i=1}^{m} \left(y_{i1}\theta_{i1} - \exp(\theta_{i1}) - \ln(y_{i1}!)\right) + \sum_{j=2}^{n_i} \left(y_{ij}{\theta_{ij}}^* - \exp({\theta_{ij}}^*) - \ln(y_{ij}!)\right),$$

where $\theta_{i1} = \ln(\lambda_{i1}) = x'_{i1}\beta$ and ${\theta_{ij}}^* = \ln({\lambda_{ij}}^*)$.

The following constraints must be satisfied in order for the constructed likelihood to be valid: (1) $\lambda_{ij} > 0, (j = 1, \ldots, n_i)$; (2) $-1 < \alpha < 1$ for $(j = 2, \ldots, n_i)$, in order to achieve a positive-definite correlation matrix; and (3) $\lambda_{ij} - \alpha\sigma_{ij}/\sigma_{ij-1}(\lambda_{ij-1}) > 0$, for $(j = 2, \ldots, n_i)$ (Guerra and Shults, 2014).

4.2.3. Likelihood Equations

To obtain maximum likelihood estimates of β and α, we need to obtain simultaneous solutions to the following estimating equations for β and α, respectively:

$$\begin{aligned} \frac{\partial \ln\left(L(\beta,\alpha)\right)}{\partial\beta} &= \sum_{i=1}^{m}\left(y_{i1}-\exp(\theta_{i1})\right)\frac{\partial\theta_{i1}}{\partial\beta}+\sum_{j=2}^{n_i}\left(y_{ij}-\exp({\theta_{ij}}^*)\right)\frac{\partial{\theta_{ij}}^*}{\partial\beta} \\ &= 0 \end{aligned} \tag{4.7}$$

and

$$\begin{aligned} \frac{\partial \ln\left(L(\beta,\alpha)\right)}{\partial\alpha} &= \sum_{i=1}^{m}\left(y_{i1}-\exp(\theta_{i1})\right)\frac{\partial\theta_{i1}}{\partial\alpha}+\sum_{j=2}^{n_i}\left(y_{ij}-\exp({\theta_{ij}}^*)\right)\frac{\partial{\theta_{ij}}^*}{\partial\alpha} \\ &= 0. \end{aligned} \tag{4.8}$$

The derivatives are given in Appendix Section C.1.

There is no explicit solution to the likelihood equations 4.7 and 4.8. We obtained solutions by maximizing the likelihood using an adaptive barrier algorithm as implemented in the `constrOptim` function in R (R Development Core Team, 2012). We applied the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimization method by Broyden, 1970, Fletcher, 1970, Goldfarb, 1970, and Shanno, 1970; Shanno and Kettler, 1970, which is implemented in `constrOptim` when the gradient is provided.

The following algorithm summarizes our estimation procedure for a particular model:

1. Choose initial estimates (starting values) of α and β. Starting values can be obtained using GEE to fit a Poisson model with an AR(1) correlation structure; however, we should check that the starting values satisfy the constraints (Section 4.2.2). If the estimates violate the constraints, change the starting values by choosing a value for α that is closer to zero or by applying Poisson regression, which is equivalent to assuming that $\alpha = 0$.

2. Obtain solutions to the likelihood equations 4.7 and 4.8 using the adaptive barrier algorithm that is implemented in the R package `constrOptim`. Refer to the Appendix C.2 for the log likelihood function and Appendix C.3 for the gradient function, both of which are implemented in the Application.

4.2.4. Asymptotic Distribution of the Estimators

If the model is correctly satisfied and standard regularity conditions are satisfied, the ML approach described here will yield estimates that are consistent and asymptotically normal. Define the vector of parameters as $\theta = (\beta, \alpha)^T$ and the maximum likelihood estimators as $\hat{\theta} = (\hat{\beta}, \hat{\alpha})^T$. The asymptotic covariance matrix of $\hat{\theta}$ is the observed information $(i(\hat{\theta}))^{-1}$, which we estimated using the inverse of the negative Hessian matrix (Appendix Section C.6).

4.3. Application

4.3.1. Doctor visits data

Here we consider an analysis of a subset of data from the German Socio-Economic Panel data (Winkelmann, 2004) (`http://www.stata-press.com/data/r13/drvisits`) that we obtained within Stata (StataCorp LP, 2013) and then exported as a comma delimited text-file for analysis in R. These data were analyzed in the Stata 13 mixed-effects reference manual (StataCorp LP, 2013) with generalized linear mixed-effects models that included subject level random intercepts. Here we compare the results of an analysis using the proposed ML approach with the results obtained using Poisson regression and GEE.

The goal of the analysis was to assess the impact of the 1997 health reform on the reduction of government expenditures. A sample of 1518 women who were employed full time in the year before or in the year after the reform was used to assess the impact on the number of doctor visits.

The outcome was the self-reported number of doctor visits in the three months prior to the interview. The main covariate of interest was the indicator of whether the interview was before the reform or after it. Additional covariate information was available on the women's age, education, marital status, self-reported health status, and the logarithm of the household income. Note that not every woman was interviewed both before and after the reform went into effect. Of the 1518 women in the dataset, 709 were interviewed both before and after the reform and the remaining 809 were interviewed only once (391 women before and 418 after the reform went into effect). This resulted in a total of 2227 observations available for the analysis.

We assumed Model 4.5 with the following linear predictor:

$$x_{ij} = \beta_0 + \beta_1 x_{1ij} + \beta_2 x_{2ij} + \beta_3 x_{3ij} + \beta_4 x_{4ij} + \beta_5 x_{5ij} + \beta_6 x_{6ij},$$

where x_{ij1} was the indicator for reform, x_{ij2} was age in years, x_{ij3} was education in years, x_{ij4} was marital status, x_{ij5} was self-reported health status, and x_{ij6} was the logarithm of household income.

We first fit the above model using Poisson regression as implemented in the `glm` function in R. This assumed that the longitudinal counts of doctor visits before and after the reform are independent, given the covariates. Therefore, we did not account for the correlation among the repeated measures of the doctor visit counts in the estimation, possibly leading to unreliable estimates of the standard errors. Table 4.1 shows the estimated regression coefficients, standard errors, test statistics, and p-values for this model. From the model, the expected change in log count of doctor visits from before to after the reform was -0.140 ($p < 0.0001$).

Next, we applied the GEE approach to the analysis of this data using the `geeglm` function in R. Table 4.1 shows the estimates and results for this model. As for Poisson regression, there was a significant effect of the reform indicating a change in the count of the number of doctor visits from before to after the reform ($\hat{\beta} = -0.123, p = 0.0200$). The estimated correlation parameter was 0.213.

When we fit the GEE model we assumed that the scale parameter ϕ is equal to one. After fitting GEE, we can assess the adequacy of this assumption by obtaining an estimate of ϕ based on the final GEE estimates of β:

$$\widehat{\phi} = \frac{1}{m} \sum_{i=1}^{m} \frac{Z_i(\widehat{\boldsymbol{\beta}})' Z_i(\widehat{\boldsymbol{\beta}})}{n_i},$$

where $Z_i(\widehat{\boldsymbol{\beta}})$ is the $n_i \times 1$ vector of Pearson residuals $z_{ij}(\widehat{\boldsymbol{\beta}})$ with $z_{ij}(\widehat{\boldsymbol{\beta}}) = \frac{y_{ij} - \widehat{\lambda_{ij}}}{\sqrt{\widehat{\lambda_{ij}}}}$. The estimated ϕ is $\hat{\phi} = 4.33$, which is much greater than 1 and is therefore suggestive of over-dispersion in the data.

Lastly, we fit the proposed ML approach using the algorithm for estimation described in Section 4.2.3. We obtained starting values for our approach using GEE, after first confirming that $\hat{\alpha}$ satisfied the necessary constraint to guarantee a valid parent distribution, which in this case was $\hat{\alpha} < 0.4518$.

Table 4.1 shows the estimated regression coefficients, standard errors, test statistics, and p-values for the ML approach. The estimated correlation parameter was 0.313 with a 95% confidence interval of (0.272, 0.354). Although not customary for longitudinal data, a likelihood ratio test of the null hypothesis $\alpha = 0$ resulted in a p-value < 0.0001. After adjusting for the correlation among the counts of doctor visits, for over-dispersion, and for the other covariates, we found that there was a significant impact of the reform on the number of doctor visits based on the ML model ($\hat{\beta}_1 = -0.113, p < 0.0001$).

Overall, the parameter estimates were similar for the proposed ML approach, GEE, and the Poisson regression. While the impact of age was similar across the approaches, it was significant in both the ML and Poisson approaches but not significant in the GEE model (ML $p = 0.0005$, GEE $p = 0.1180$, and Poisson $p = 0.0008$). Similarly, the logarithm of household income was significant in both the ML and Poisson approaches but not significant in the GEE model (ML $p < 0.0001$, GEE $p = 0.0810$, and Poisson $p < 0.0001$).

With estimates of the log-likelihood for Poisson regression and the proposed ML approach, it was possible to calculate the AIC and BIC criteria as measures of the relative quality of the models for this set of data. Both BIC and AIC incorporate a penalty term for the number of parameters used in the model because it is possible to increase the numerical value of the likelihood solely by including additional parameters in the model, which may result in over-fitting the model to the data. This penalty term is larger in the BIC as compared to the AIC.

The AIC was computed as 2 times the degrees of freedom represented by the number of parameters in this model minus 2 times the estimated log-likelihood. The BIC was computed as the number of parameters estimated times the logarithm of the sample size minus 2 times the estimated log-likelihood.

For the Poisson regression model, the AIC and BIC values were 11899 and 1196, which were both greater than the AIC and BIC values for the ML approach (AIC = 11707 and BIC = 11750), which indicates that the ML approach had improved model fit over Poisson regression. R code for this analysis is provided in the Appendix Section C.7.

4.3.2. Epilepsy seizure data

Here we implement the proposed ML method and GEE for analysis of the epilepsy seizure data (Farewell and Farewell, 2012) (Thall and Vail, 1990). We do not demonstrate the application of Poisson regression as we did in the previous section. However, results for Poisson regression (not shown) all confirmed the selection of the more general model assumed by the proposed ML approach.

We assumed Model 4.5 with the following linear predictor:

$$x'_{ij}\beta = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3} + \beta_4 x_{ij4}, \tag{4.9}$$

where x_{ij1} represents an indicator for treatment, x_{ij2} represents baseline seizure count (number of seizures in the 3 month time period prior to the start of the study), x_{ij3} represents subject age in years, and x_{ij4} represents two-week time period (coded as 1,2,3,4). We initially included a time period by treatment interaction term, but the interaction term was not significant for the proposed approach or for GEE (all p-values > 0.05); we therefore initially focused on the simpler model 4.9 for this demonstration.

Table 4.2 shows the sample mean and variance of seizure counts at baseline and the four subsequent two-week periods (denoted as Y1 through Y4) for the placebo and drug groups for the seizure counts; it also displays the sample mean and variance of age at baseline. From the table, the sample variance for the outcome variables, Y1 through Y4, were greater than their respective means, which suggested that there was over-dispersion the seizure counts.

Table 4.3 shows the results of the analysis. The estimates were similar for the proposed ML method and GEE. The estimate of treatment was negative for both approaches, which suggested that the number of seizures was lower for subjects in the treatment group. However, treatment only differed significantly from 0 for the proposed ML approach ($p = 0.0127$ for ML versus $p = 0.3014$ for GEE). In addition, time period only differed significantly from 0 for the proposed ML approach ($p = 0.0031$ for ML versus $p = 0.0580$ for GEE).

The likelihood ratio test of the hypothesis that the regression parameter for time period is 0 also suggested that time period should be retained in the model for the proposed ML approach (p =

0.0030). However, since the GEE analysis suggested that time period might not be important, we removed period from the GEE model. For comparison, we also removed period for the proposed ML approach. The analysis results are shown in Table 4.4. Perhaps the most interesting feature of these new results was that treatment no longer differed significantly from 0 for the proposed ML approach (and was again not significant for GEE). This analysis demonstrated that removal of a significant variable (time period) from the model for the proposed ML approach resulted in the treatment effect no longer being significant.

A treatment effect whose significance depended on the inclusion of an additional variable in the model should be assessed carefully. We therefore compared the AIC and BIC for the larger model that included time period with the smaller model that does not include period. As shown in the Tables, the AIC and BIC values were both smaller for the larger model. The respective AIC and BIC values were 1566 and 1579 for the larger model, versus 1573 and 1583 for the smaller model. The AIC and BIC values indicated that the fit was superior for the larger model, which lent additional support for the larger model with its significant treatment effects.

4.4. Simulation Studies

In the previous section we identified significant treatment effects for the proposed ML approach that were not observed for GEE. Since the results depended on choice of approach, it was of interest to compare the performance of the methods for finite samples. We therefore performed simulations to assess the properties of the estimators of α and β for the proposed ML approach and GEE.

4.4.1. Set-up

We compared the performance of the ML and GEE estimators for the final GEE model that was implemented in the previous section; the linear predictor for this model was:

$$x'_{ij}\beta = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3}, \tag{4.10}$$

where the x_{ijk} were defined in the previous section. The results shown here are based on $R = 1000$ simulation runs, β = $(0.4467, -0.1659, 0.0232, 0.0258)'$ equal group sizes $m/2$, and n_i = 4 measurements per subject. For this scenario, the correlation must satisfy the following constraints (see

Section 4.2.2) to ensure the existence of a valid parent distribution:

$$\alpha < 0.707.$$

We specified values of $\alpha \in \{0.2, 0.4, 0.6, 0.7\}$.

Covariates were simulated based on the observed data in the previous section. Treatment was specified as present (equal to 1) for one group and as absent (equal to 0) for the other group. Baseline seizure count was simulated from a Poisson distribution with a random seed and mean $= 31.22$ based on the mean baseline seizure counts from the epilepsy data. Similarly, age was simulated from a normal distribution based on the epilepsy data for which the minimum age was 18, mean was 28.3, and the standard deviation was 6.261. Simulated age values below 18 were discarded and the next simulated age value was assigned. Age was then rounded to a whole number, as it was recorded in the epilepsy data.

The approach proposed by Guerra and Shults, 2014 was used to simulate the correlated Poisson seizure counts with specified means, over-dispersion, and AR(1) correlation structure.

4.4.2. Assessments

We wrote code in R to evaluate percent bias, small sample efficiency, and 95% coverage probabilities using the observed information matrix. Details on how these were calculated are provided below.

Let θ represent a parameter of interest for the evaluation and $\hat{\theta}$ represent the estimator. The mean square error (MSE) for estimator $\hat{\theta}$ is defined as

$$\frac{1}{R}\sum_{i=1}^{R}(\theta - \hat{\theta}_i)^2,$$

where θ is the true value. The percent bias for estimator $\hat{\theta}$ is defined as

$$\left\{\frac{1}{R}\sum_{i=1}^{R}(\theta - \hat{\theta}_i)/\theta\right\} * 100.$$

Lastly, to evaluate the coverage probabilities, a 95% confidence interval was computed for each

parameter estimate within each simulation run. The coverage probabilities represent the proportion of the R simulation runs in which the true parameter fell within the 95% confidence bounds. GEE coverage probabilities were computed similarly using the naïve variance estimates obtained from `geeglm` in R.

4.4.3. Results

Table 4.5 displays the MSE and Table 4.6 displays the percent bias for the simulations. For the ML method, the MSE for $\hat{\beta}$ and $\hat{\alpha}$ and the percent bias for $\hat{\alpha}$ decreased as m increased.

As compared to GEE, the ML approach had lower MSE and percent bias for all sample sizes for $\hat{\alpha}$. For $\hat{\beta}$, the percent bias was similar for ML and GEE; however, the MSE was slightly smaller for ML than for GEE. For scenarios with high correlation ($\alpha = 0.6$ or 0.7), the intercept and treatment estimates, $\hat{\beta}_0$ and $\hat{\beta}_1$, had smaller MSE and percent bias for the proposed ML approach than for GEE, for all samples sizes.

Table 4.7 then displays the estimated coverage probabilities. With respect to $\widehat{\beta}$, the coverage probabilities were similar for the ML and GEE approach and were close to the nominal 95% level. With respect to $\widehat{\alpha}$, the ML approach model-based coverage probabilities were close to the nominal 95%, which outperformed the GEE approach, whose model-based coverage probabilities were below the nominal 95% level. Coverage probabilities for α were better for the ML based approach than GEE across all sample sizes and correlations ($\alpha = 0.2, 0.4, 0.6, 0.7$).

4.5. Discussion

We proposed an ML approach for analysis of equally spaced longitudinal count data that accounts for intra-subject correlation of measurements and over-dispersion. Our application of the ML and GEE approaches demonstrated significant treatment differences observed for some models for the ML approach but not for GEE. The availability of the AIC and BIC criteria for the ML approach was useful for selecting between nested models, when the significance of the treatment effects depended on the inclusion of time in the model.

Our simulations demonstrated that the ML approach was similar to or slightly outperformed GEE with respect to MSE, bias, and coverage probabilities, especially for higher values of the correlation

(for $\hat{\beta}$).

That the ML approach outperformed GEE for larger values of the correlation was not surprising. We assumed over-dispersion that was induced by α and that was greater for larger values of α. For $\alpha = 0$ the assumed models for the marginal means and correlations would have been identical for the ML approach and GEE. That the differences between the two approaches were largest for larger values for the correlation was therefore to be expected.

Future work might be undertaken to compare the proposed approach with implementation of the generalized Poisson distribution (Famoye, Okafor, and Adamu, 2011) that can be used to assess over- and under-dispersion in correlated count data but that does not implement the usual model for the marginal mean in Poisson regression. Extensions for unequally spaced data will also be useful.

Our approach is also valid when data are missing at random, while the GEE approach assumes the missing data mechanism to be missing completely at random which is more restrictive (Liang and Zeger, 1986). Future work will include assessing the loss in efficiency for GEE relative to our approach, when the data are missing at random. In addition, it will be useful to extend our approach to allow for application of other correlation structures such as the Markov structure. The Markov structure generalizes the AR(1) structure to take the spacing of measurements into account and is therefore a plausible structure for data that are unequally spaced in time.

Table 4.1: Estimated parameters from the ML, GEE, and Poisson models in the analysis of the doctor visits data.

ML Approach ($AIC = 11707; BIC = 11750$)				
Coefficients:				
Parameter	Estimate	Std.err	Wald	$Pr(>\|W\|)$
(Intercept)	-0.461	0.2811	2.69	0.1010
Reform	-0.113	0.0241	21.99	< 0.0001
Age	0.005	0.0014	12.22	0.0005
Education	-0.008	0.0064	1.54	0.2150
Marital Status	0.026	0.0294	0.75	0.3860
Health Status	1.100	0.0313	1238.28	< 0.0001
Log Income	0.150	0.0376	15.83	< 0.0001
Correlation Parameters:				
Parameter	Estimate	Std.err		
alpha	0.313	0.0208		
GEE Approach				
Coefficients:				
Parameter	Estimate	Std.err	Wald	$Pr(>\|W\|)$
(Intercept)	-0.381	0.5767	0.44	0.5080
Reform	-0.123	0.0530	5.40	0.0200
Age	0.005	0.0033	2.44	0.1180
Education	-0.009	0.0118	0.61	0.4350
Marital Status	0.038	0.0698	0.30	0.5820
Health Status	1.105	0.0873	160.23	< 0.0001
Log Income	0.139	0.0798	3.05	0.0810
Correlation Parameters:				
Parameter	Estimate	Std.err		
alpha	0.213	0.0238		
Poisson Regression ($AIC = 11899; BIC = 11936$)				
Coefficients:				
Parameter	Estimate	Std.err	Wald	$Pr(>\|W\|)$
(Intercept)	-0.414	0.2691	-1.54	0.1242
Reform	-0.140	0.0266	-5.28	< 0.0001
Age	0.004	0.0013	3.35	0.0008
Education	-0.011	0.0060	-1.78	0.0743
Marital Status	0.041	0.0278	1.49	0.1375
Health Status	1.133	0.0303	37.40	< 0.0001
Log Income	0.149	0.0361	4.14	< 0.0001

Table 4.2: Mean and variance for the placebo and treatment groups.

Variable	Placebo† (n=28)	Drug† (n=31)	Total† (n=59)
Y1	9.86 (102.8)	8.58 (332.7)	8.95 (220.2)
Y2	8.29 (66.6)	8.42 (140.7)	8.36 (103.8)
Y3	8.79 (215.2)	8.13 (192.9)	8.44 (200.2)
Y4	7.96 (58.2)	6.71 (126.8)	7.31 (93.1)
Baseline	30.79 (681.2)	31.61 (782.9)	31.22 (722.5)
Age	29.00 (36.0)	27.74 (43.6)	28.34 (39.7)

† Values in the table represent the mean (variance).

Table 4.3: Estimated parameters from the GEE and ML approaches for analysis of the epilepsy data when Period is included in the models.

ML Approach ($AIC = 1566; BIC = 1579$)				
Coefficients:				
Parameter	Estimate	Std.err	Wald	$Pr(>\|W\|)$
(Intercept)	0.657	0.1957	11.26	0.0008
Treatment	-0.166	0.0667	6.21	0.0127
Baseline	0.023	0.0007	1111.76	< 0.0001
Age	0.024	0.0056	17.94	< 0.0001
Period	-0.064	0.0215	8.74	0.0031
Correlation Parameters:				
Parameter	Estimate	Std.err		
alpha	0.416	0.0334		
GEE Approach				
Coefficients:				
Parameter	Estimate	Std.err	Wald	$Pr(>\|W\|)$
(Intercept)	0.585	0.3491	2.81	0.0936
Treatment	-0.164	0.1589	1.07	0.3014
Baseline	0.023	0.0012	350.97	< 0.0001
Age	0.026	0.0118	4.95	0.0261
Period	-0.064	0.0340	3.59	0.0580
Correlation Parameters:				
Parameter	Estimate	Std.err		
alpha	0.551	0.0656		

Table 4.4: Estimated parameters from the GEE and ML approaches for analysis of the epilepsy data when Period is not included in the models.

ML Approach ($AIC = 1572.99; BIC = 1583.39$)				
Coefficients:				
Parameter	Estimate	Std.err	Wald	$Pr(>\|W\|)$
(Intercept)	0.5072	0.3829	1.75	0.1853
Treatment	-0.1673	0.1521	1.21	0.2713
Baseline	0.0232	0.0012	351.45	$< .0001$
Age	0.0238	0.0127	3.51	0.0611
Correlation Parameters:				
Parameter	Estimate	Std.err		
alpha	0.423	0.0668		
GEE Approach				
Coefficients:				
Parameter	Estimate	Std.err	Wald	$Pr(>\|W\|)$
(Intercept)	0.4467	0.3621	1.52	0.2174
Treatment	-0.1659	0.1593	1.09	0.2977
Baseline	0.0232	0.0012	353.32	$< .0001$
Age	0.0258	0.0117	4.86	0.0275
Correlation Parameters:				
Parameter	Estimate	Std.err		
alpha	0.544	0.0639		

Table 4.5: Small sample efficiencies for evaluating the AR(1) correlation structure for varying values of α and sample size per group.

			Method: ML Mean squared error						Method: GEE Mean squared error				
m	α	R*	$\hat{\beta}_0$	$\hat{\beta}_1[1]$	$\hat{\beta}_2[2]$	$\hat{\beta}_3[2]$	$\hat{\alpha}[1]$	R*	$\hat{\beta}_0$	$\hat{\beta}_1[1]$	$\hat{\beta}_2[2]$	$\hat{\beta}_3[2]$	$\hat{\alpha}[1]$
60	0.2	1000	0.056	0.355	0.297	0.291	0.609	1000	0.057	0.355	0.300	0.290	0.668
	0.4	1000	0.088	0.503	0.427	0.445	0.505	1000	0.089	0.516	0.427	0.450	0.701
	0.6	1000	0.127	0.803	0.642	0.619	0.308	1000	0.137	0.852	0.703	0.653	0.571
	0.7	998	0.132	0.908	0.716	0.656	0.171	1000	0.160	1.133	0.883	0.795	0.424
120	0.2	1000	0.029	0.176	0.138	0.137	0.305	1000	0.029	0.176	0.139	0.138	0.340
	0.4	1000	0.040	0.254	0.203	0.194	0.236	1000	0.040	0.260	0.204	0.198	0.334
	0.6	1000	0.054	0.381	0.291	0.294	0.124	1000	0.062	0.415	0.327	0.325	0.240
	0.7	1000	0.067	0.489	0.349	0.325	0.067	1000	0.083	0.595	0.435	0.402	0.178
300	0.2	1000	0.010	0.071	0.057	0.054	0.111	1000	0.010	0.072	0.058	0.054	0.128
	0.4	1000	0.016	0.101	0.084	0.078	0.080	1000	0.017	0.103	0.085	0.079	0.124
	0.6	1000	0.025	0.153	0.121	0.118	0.047	1000	0.027	0.162	0.132	0.129	0.093
	0.7	1000	0.029	0.174	0.144	0.140	0.023	1000	0.036	0.211	0.182	0.176	0.066

Note: The true correlation structure is AR(1). There are equal sample sizes of $\frac{m}{2}$ per group and $\beta = (\beta_0, \beta_{drug}, \beta_{baseline}, \beta_{age})' = (0.4467, -0.1659, 0.0232, 0.0258)'$; [1]True value by a factor of 10^2; [2]True value by a factor of 10^4;

Table 4.6: Percent bias for evaluating the AR(1) correlation structure for varying values of α and sample size per group.

			Method: ML Percent bias						Method: GEE Percent bias				
m	α	R*	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\alpha}$	R*	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\alpha}$
60	0.2	1000	2.57	0.53	-0.61	-0.53	9.41	1000	2.48	0.54	-0.60	-0.49	10.94
	0.4	1000	6.33	-0.42	-1.15	-2.31	5.15	1000	6.26	-0.34	-1.10	-2.29	6.06
	0.6	1000	1.95	0.05	-0.95	0.27	2.65	1000	1.88	0.64	-0.90	0.45	4.86
	0.7	998	-2.21	2.71	0.72	1.21	0.69	1000	0.60	1.87	-0.28	0.84	4.60
120	0.2	1000	-0.04	0.14	0.07	0.20	5.30	1000	-0.22	0.07	0.13	0.24	6.19
	0.4	1000	2.25	-0.52	-0.57	-0.65	2.74	1000	1.95	-0.51	-0.40	-0.64	2.89
	0.6	1000	0.43	-0.79	0.08	-0.13	1.39	1000	0.16	-1.05	0.34	-0.21	2.18
	0.7	1000	2.00	0.15	-0.01	-1.04	0.17	1000	1.74	-0.22	0.14	-0.83	2.55
300	0.2	1000	0.68	-0.18	-0.57	0.22	2.83	1000	0.65	-0.23	-0.57	0.23	2.87
	0.4	1000	0.85	-0.29	-0.16	-0.25	1.31	1000	0.72	-0.32	-0.15	-0.18	0.98
	0.6	1000	1.91	-0.38	-0.30	-0.75	0.53	1000	2.03	-0.33	-0.23	-0.88	0.86
	0.7	1000	1.47	-0.29	-0.14	-0.63	-0.03	1000	2.83	-0.20	-0.58	-0.91	1.64

Note: The true correlation structure is AR(1). There are equal sample sizes of $\frac{m}{2}$ per group and $\beta = (\beta_0, \beta_{drug}, \beta_{baseline}, \beta_{age})' = (0.4467, -0.1659, 0.0232, 0.0258)'$; [1]True value by a factor of 10^2; [2]True value by a factor of 10^4;

Table 4.7: Coverage probabilities for the ML and GEE approaches with the AR(1) correlation structure for varying values of α and sample size per group.

				Coverage Probability				
m	α	Method	R	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\alpha}$
60	0.2	ML	1000	94.7	95.2	95.5	95.5	93.8
		GEE	1000	94.4	95.0	94.8	95.1	91.1
	0.4	ML	1000	93.8	94.6	95.9	93.0	94.6
		GEE	1000	93.2	94.3	95.5	92.7	86.1
	0.6	ML	1000	93.8	93.9	94.3	94.0	93.4
		GEE	1000	94.1	93.6	95.1	93.1	83.2
	0.7	ML	998	95.4	95.3	95.4	95.5	92.3
		GEE	1000	95.0	94.9	94.0	95.7	84.6
120	0.2	ML	1000	94.7	95.2	95.2	94.8	92.9
		GEE	1000	94.2	95.1	94.9	94.5	91.3
	0.4	ML	1000	95.1	96.1	95.6	94.7	95.1
		GEE	1000	95.2	96.0	95.5	94.5	85.4
	0.6	ML	1000	95.9	94.5	95.3	94.9	95.5
		GEE	1000	95.5	95.5	95.5	94.9	84.5
	0.7	ML	1000	95.3	94.2	94.7	96.2	92.9
		GEE	1000	95.3	94.2	95.0	95.9	87.2
300	0.2	ML	1000	95.2	95.0	94.7	94.7	94.5
		GEE	1000	95.6	95.3	94.8	94.6	91.5
	0.4	ML	1000	93.5	95.4	94.2	93.9	96.5
		GEE	1000	93.7	96.0	94.9	94.3	86.2
	0.6	ML	1000	93.2	95.4	94.9	94.0	95.2
		GEE	1000	93.8	95.6	94.6	94.9	85.9
	0.7	ML	1000	94.5	95.1	94.1	94.4	92.4
		GEE	1000	94.8	95.9	94.6	94.8	88.0

Note: The true correlation structure is AR(1). There are equal sample sizes of $\frac{m}{2}$ per group and $\beta = (\beta_0, \beta_{drug}, \beta_{baseline}, \beta_{age})' = (0.4467, -0.1659, 0.0232, 0.0258)'$

CHAPTER 5

DISCUSSION

The objective of this dissertation was to develop a new and potentially clinically useful methodology for time-conditional survival probability analyzed under the nonparametric and the parametric frameworks. Further, we addressed the problem of correlated longitudinal data with over-dispersion by developing a maximum likelihood analysis of discrete count data with over-dispersed marginal distributions relative to the Poisson distributions and an induced AR(1) correlation structure.

In Chapter 2, we developed the asymptotic distribution for the estimated log time-conditional survival probabilities. Weighted least squares methodology was used to develop a hypothesis testing framework to address clinically relevant questions of interest, e.g. a multivariate omnibus test of pairwise differences in time-conditional survival probabilities. We also fit regression models for the log time-conditional survival probabilities as a function of time survived to explore the relationship (quadratic, linear, global mean) between the probabilities and time survived as well as covariates. In simulations, the test statistic for the global mean test had good statistical properties for samples of size 200 or greater but the sample size necessary for good statistical properties increased as the percentage of uniform random censoring increased. We found that the sample size necessary to achieve adequate power for this test also increased with increasing percentage of uniform random censoring.

Regression models were developed for the profile of log time-conditional survival probabilities as a function of time survived after diagnosis and included adjustment for covariates. This modeling approach is an improvement on profile-based methods, where disparate profiles are created based on the covariate patterns, as it allows for the evaluation of the importance of profiles and factors. Using our methodology, we explored time-conditional survival probability profiles developed using a dataset of patients with malignant melanoma and their survival probabilities. Our formal statistical methods allow researchers to identify when time-conditional estimates change in statistically significant ways.

In Chapter 3, we developed methods for parametric time-conditional survival probability by incorporating multiple covariates, including continuous variables. We first described the general approach

for including covariates under the parametric survival regression model framework and then implemented this for the Weibull regression model and the Logistic-Weibull cure model. Further, we presented methodology to evaluate time-conditional survival as a function of both categorical and continuous covariates at continuous times after diagnosis for any possible value of future survival beyond the current survival time.

When comparing methods in parametric and nonparametric survival, the choice of approach depends on a variety of factors. It is well known that misspecification of the true underlying parametric distribution can lead to a loss of power as compared with the nonparametric empirical survival distribution (Hutton and Monaghan, 2002). Alternately, the product-limit method of estimating nonparametric survival distributions only allows for the consideration of categorical covariates by estimating survival probabilities based on the resulting covariate patterns. As a result, the full sample size is reduced to the number of subjects that fall into each covariate pattern for estimation of the survival probabilities. Further, it has been shown that with a limited follow-up of 5 years the nonparametric log-rank test has greater statistical power to reject the null hypothesis when it is false than its parametric counterpart (Gamel and Vogel, 1997).

This advantage declines as follow-up time increases (Gamel and Vogel, 1997), which is an important consideration for time-conditional survival probability estimation. Although nonparametric methods may be able to distinguish between groups, they may not help the researcher understand the underlying relationships between covariates and the difference in survival. However, the cure model can distinguish covariate effects on the survival time-to-an-event rather than on the proportion cured of the event.

Incorporating this detailed covariate information adds an element of personalization to the estimates by allowing clinicians and investigators to customize the estimate of interest to the characteristics relevant to their patients and research subjects. In our example, patients with melanoma would not need to be stratified into groups based on their age. Instead, the exact age could be used to estimate time-conditional survival probability under the parametric approach. The Weibull distribution is used in Chapter 3 to illustrate the methods as applied to SEER esophageal cancer data (Weibull regression model with covariates) and to SEER melanoma data (Logistic-Weibull cure model). While the nonparametric approach includes an overall test of significance in addition to pairwise comparisons, the parametric approach allows for the inclusion of this variable in its continuous form to

address the preference of some researchers to analyze continuous variables in continuous rather than categorical form (Bennette and Vickers, 2012; Greenland, 1995).

As we explored and developed the statistical framework around time-conditional survival probabilities as they applied to lifetime data, many interesting questions arose. An alternative perspective on this problem, which is of future interest, is the reversal of the statement of the problem. In this dissertation we assumed a fixed choice for Δ while varying a. The reverse question can be evaluated by fixing a choice of time alive after diagnosis, a, and varying Δ. For example, this would allow the researcher to get 1-, 2-, and 3-year time-conditional survival probability estimates given survival beyond 3 years. Future research should further assess how to determine the Δ for Δ-year time-conditional survival probability estimation and how to determine the appropriate time survived (a) to be evaluated. Some of our initial exploration in this area suggests that identifiability issues may arise when trying to invert the covariance matrix for a profile of time-conditional survival probability estimates obtained from a constant portion of the survival curve due to few events. Researchers must also be cautious when computing estimates of survival probability based on sparse data as the results may be misleading. Such situations may result in larger variances and will directly impact the calculation of the test statistic, which will lead to decreased power to reject the null hypothesis.

In Chapter 4, we considered data from longitudinal studies where Poisson count outcomes are measured repeatedly on subjects over time. Our objective was to fit a Poisson regression model to relate the expected value of the outcomes with covariates, while also adjusting for over-dispersion and the intra-subject correlation of measurements.

We presented a maximum-likelihood based method for analysis of longitudinal count data. Key assumptions of our approach were the first-order Markov property; linearity of the conditional expectations; conditional distributions that were Poisson; and constant adjacent correlation of measurements. These assumptions resulted in marginal distributions with the usual marginal means for Poisson regression but with over-dispersion owing to variances that were inflated relative to the Poisson distribution (Guerra and Shults, 2014; Guerra et al., 2012). The over-dispersion was induced by the adjacent correlations, so that higher values for the correlation resulted in more over-dispersion, while zero correlation yielded the usual model for Poisson regression. The assumptions also induced an AR(1) correlation structure for the measurements pertaining to each subject. Our approach could therefore be viewed as an extension of traditional Poisson regression

for over-dispersed count data with an AR(1) correlation structure.

We provided an estimation algorithm and developed software in R to implement the algorithm that we demonstrated in two analyses. We also performed simulations to compare our method with GEE. Our simulations indicated that our approach had better small sample efficiency than GEE for all simulation scenarios, and especially for higher values of the correlation.

An additional benefit of a maximum likelihood based approach relative to a semi-parametric approach like GEE includes access to the maximized log-likelihood that can be used to conduct likelihood ratio tests and to obtain measures such as the AIC and BIC that can be used to compare nested models. If the model is correctly specified, the maximum likelihood estimators will also have smallest asymptotic variance amongst a large class of estimators. Our maximum likelihood approach is also valid when data are missing at random, while the GEE approach requires the more restrictive missing completely at random assumption regarding the missing data mechanism (Liang and Zeger, 1986).

To encourage the use of the methodology for longitudinal count data by others, we plan to develop an R package that will be based on the R code that we developed for this dissertation. Future work is planned to translate this software to other software packages, including SAS and Stata. Extensions to unequally spaced data and to other distributions might also be considered.

APPENDIX A

CHAPTER 1

A.1. Estimated Variance of $\widehat{CS}(b \mid a)$

Define the estimated nonparametric time-conditional survival probability, $\widehat{CS}(b \mid a)$, as in Equation 2.2 given by

$$\widehat{CS}(b \mid a) = \frac{\hat{S}(b)}{\hat{S}(a)} = \frac{\prod_{j:t_{(j)} \leq b} (1 - \hat{\pi}_j)}{\prod_{j:t_{(j)} \leq a} (1 - \hat{\pi}_j)} = \prod_{j:a < t_{(j)} \leq b} (1 - \hat{\pi}_j) \,.$$

In this Appendix, we derive the formula currently used in the literature to estimate symmetric confidence bands for $\widehat{CS}(b \mid a)$.

Applying the natural logarithm transformation, $\log \widehat{CS}(b \mid a) = \sum_{j:a<t_{(j)} \leq b} (1 - \hat{\pi}_j)$. By the δ-method,

$$\begin{aligned} Var\left(\log \widehat{CS}(b \mid a)\right) &\approx \sum_{j:a<t_{(j)} \leq b} \left(\frac{\partial log(1-\pi_j)}{\partial \pi_j}\right)^2 \left(Var(\hat{\pi}_j)\right) \\ &= \sum_{j:a<t_{(j)} \leq b} \left(\frac{1}{1-\pi_j}\right)^2 \left(\frac{\pi_j(1-\pi_j)}{n_j}\right) \\ &= \sum_{j:a<t_{(j)} \leq b} \frac{\pi_j}{n_j(1-\pi_j)}. \end{aligned}$$

Using the MLE for π_j, $\hat{\pi}_j = \frac{d_j}{n_j}$, an estimator of $Var\left(\log \widehat{CS}(b \mid a)\right)$ is given by

$$\widehat{Var}(\log \widehat{CS}(b \mid a)) = \sum_{j:a<t_{(j)} \leq b} \frac{d_j}{n_j(n_j - dJ)}.$$

Using the δ-method and the relationship $\widehat{CS}(b \mid a) = \exp[\log \widehat{CS}(b \mid a)]$, the variance for the estimator is given by

$$\begin{aligned} Var(\widehat{CS}(b \mid a)) &\approx \left(\frac{\partial \exp[\log \widehat{CS}(b \mid a)]}{\partial \log \widehat{CS}(b \mid a)}\right)^2 V[\log \widehat{CS}(b \mid a)] \\ &= \left(\exp[\log \widehat{CS}(b \mid a)]\right)^2 \sum_{j:a<t_{(j)} \leq b} \frac{\pi_j}{n_j(1-\pi_j)}, \end{aligned}$$

and is a variation of Greenwood's formula (Greenwood, 1926). This quantity is consistently esti-

mated by

$$\widehat{Var}\left(\widehat{CS}(b \mid a)\right) = \left[\widehat{CS}(b \mid a)\right]^2 \sum_{j:a<t_{(j)}\leq b} \frac{d_j}{n_j(n_j - d_j)},$$

and provides a large sample variance of the estimator of time-conditional survival probability. This estimated variance can be used to provide symmetric confidence bands for $\widehat{CS}(b \mid a)$.

A.2. Estimated Covariance of $CS_1(b_1 \mid a_1)$ and $CS_2(b_2 \mid a_2)$

In this Appendix, we derive the covariance of two time-conditional survival probabilities. First, consider the covariance between two log time-conditional survival probabilities given by

$$\log\left[\widehat{CS}_1(b_1 \mid a_1)\right] = \sum_{j:a_1<t_{(j)}\leq b_1} log(1-\hat{\pi}_j),$$

and

$$\log\left[\widehat{CS}_2(b_2 \mid a_2)\right] = \sum_{j:a_2<t_{(j)}\leq b_2} log(1-\hat{\pi}_j).$$

We consider two cases for the relationships between $a_1, a_2, b_1,$ and b_2. The first case is for two non-overlapping log time-conditional survival probabilities and the second is for two overlapping log time-conditional survival probabilities.

Case 1. For two non-overlapping log time-conditional survival probability estimators, define fixed event times a_1, b_1, a_2, b_2 such that $0 \leq a_1 < b_1 < a_2 < b_2 \leq t_{(J)}$. Then, the covariance of these estimators is given by

$$\begin{aligned}
&Cov\Big(\log\Big[\widehat{CS}_1(b_1 \mid a_1)\Big], \log\Big[\widehat{CS}_2(b_2 \mid a_2)\Big]\Big) \\
&\quad = Cov\Big(\sum_{j:a_1<t_{(j)}\leq b_1} \log(1-\hat{\pi}_j), \sum_{j:a_2<t_{(j)}\leq b_2} \log(1-\hat{\pi}_j)\Big) \\
&\quad = Cov\Big(\log(1-\hat{\pi}_{a_1}) + \log(1-\hat{\pi}_{a_1+1}) + \cdots + \log(1-\hat{\pi}_{b_1}), \\
&\qquad \log(1-\hat{\pi}_{a_2}) + \log(1-\hat{\pi}_{a_2+1}) + \cdots + \log(1-\hat{\pi}_{b_2})\Big) \\
&\quad = Cov(\log(1-\hat{\pi}_{a_1}), \log(1-\hat{\pi}_{a_2})) + Cov(\log(1-\hat{\pi}_{a_1}), \log(1-\hat{\pi}_{a_1+1})) \\
&\qquad + \cdots + Cov(\log(1-\hat{\pi}_{b_1}), \log(1-\hat{\pi}_{b_2})) \\
&\quad = 0,
\end{aligned}$$

since $Cov(\hat{\pi}_j, \hat{\pi}_k) = 0, \forall j \neq k$ (Lachin, 2000). That is, if (a_1, b_1) and (a_2, b_2) are chosen such that these log time-conditional survival proportions sum over non-overlapping intervals, then the time-conditional survival proportions obtained from these fixed event times are uncorrelated.

Case 2. For two overlapping log time-conditional survival proportions, consider fixed event times a_1, b_1, a_2, b_2 such that $0 \leq a_1 \leq a_2 \leq b_1 \leq b_2 \leq t_{(J)}$. The covariance between these estimators is

given by

$$
\begin{aligned}
&Cov\Big(\log\Big[\widehat{CS}_1(b_1 \mid a_1)\Big], \log\Big[\widehat{CS}_2(b_2 \mid a_2)\Big]\Big) \\
&\quad = Cov\Big(\sum_{j:a_1<t_{(j)}\le b_1} \log(1-\hat{\pi}_j), \sum_{j:a_2<t_{(j)}\le b_2} \log(1-\hat{\pi}_j)\Big) \\
&\quad = Cov\Big(\sum_{j:a_1<t_{(j)}\le a_2} \log(1-\hat{\pi}_j) + \sum_{j:a_2<t_{(j)}\le b_1} \log(1-\hat{\pi}_j), \\
&\qquad \sum_{j:a_2<t_{(j)}\le b_1} \log(1-\hat{\pi}_j) + \sum_{j:b_1<t_{(j)}\le b_2} \log(1-\hat{\pi}_j)\Big) \\
&\quad = Cov\Big(\sum_{j:a_1<t_{(j)}\le a_2} \log(1-\hat{\pi}_j), \sum_{j:a_2<t_{(j)}\le b_1} \log(1-\hat{\pi}_j)\Big) \\
&\qquad + Cov\Big(\sum_{j:a_1<t_{(j)}\le a_2} \log(1-\hat{\pi}_j), \sum_{j:b_1<t_{(j)}\le b_2} \log(1-\hat{\pi}_j)\Big) \\
&\qquad + Cov\Big(\sum_{j:a_2<t_{(j)}\le b_1} \log(1-\hat{\pi}_j), \sum_{j:a_2<t_{(j)}\le b_1} \log(1-\hat{\pi}_j)\Big) \\
&\qquad + Cov\Big(\sum_{j:a_2<t_{(j)}\le b_1} \log(1-\hat{\pi}_j), \sum_{j:b_1<t_{(j)}\le b_2} \log(1-\hat{\pi}_j)\Big) \\
&\quad = Var\Big(\sum_{j:a_2<t_{(j)}\le b_1} \log(1-\hat{\pi}_j)\Big) \\
&\quad = Var(\log\widehat{CS}(b_1 \mid a_2)) = \sum_{j:a_2<t_{(j)}\le b_1} \frac{\pi_j}{n_j(1-\pi_j)},
\end{aligned}
$$

and this is estimated by

$$
\widehat{Var}(\log\widehat{CS}(b_1 \mid a_2)) = \sum_{j:a_2<t_{(j)}\le b_1} \frac{d_j}{n_j(n_j-d_j)}.
$$

This is the covariance formula given by Equation 2.11 in Section 2.3.

When considering the covariance between two estimators of time-conditional survival probabilities, we use the δ-method and the relationship,

$$
\widehat{CS}(b_1 \mid a_2) = \exp[\log\widehat{CS}(b_1 \mid a_2)],
$$

to obtain the variance given by

$$\begin{aligned} Var(\widehat{CS}(b_1 \mid a_2)) &\cong \left(\frac{\partial \exp[\log \widehat{CS}(b_1 \mid a_2)]}{\partial \log(\widehat{CS}(b_1 \mid a_2)} \right)^2 Var[\log \widehat{CS}(b_1 \mid a_2)] \\ &= \left(\exp[\log \widehat{CS}(b_1 \mid a_2)] \right)^2 \sum_{j:a_2 < t_{(j)} \leq b_1} \frac{\pi_j}{n_j(1-\pi_j)}, \end{aligned}$$

which is estimated by

$$\widehat{Var}(\widehat{CS}(b_1 \mid a_2)) = \left[\widehat{CS}(b_1 \mid a_2) \right]^2 \sum_{j:a_2 < t_{(j)} \leq b_1} \frac{d_j}{n_j(n_j - d_j)}$$

as in Equation 2.3.

A.3. Estimated Mean and Variance of $\log \widehat{CS}$

The Central Limit Theorem, the Law of Large Numbers, and Slutsky's Theorem are used to derive the large sample distribution of $\log \widehat{CS}$. Define a sample of observations (t_i, δ_i), $i = 1, \ldots, n$, observed at J distinct times, $t_{(1)} < t_{(2)} < \cdots < t_{(J)}$ where the underlying survival distribution is a step function that has points of discontinuity at these event times. Then the likelihood is given by

$$L(\pi_1, \pi_2, \ldots, \pi_J) \propto \prod_{j=1}^{J} \pi_j^{d_j} (1 - \pi_j)^{(n_j - d_j)},$$

where π_j is the conditional probability of an event at $t_{(j)}$, n_j is the number of subjects at risk or still under observation at time $t_{(j)}$, and d_j is the number of events observed at time $t_{(j)}$ among the n_j subjects at risk at time $t_{(j)}$. Information on censoring is accounted for by defining w_j as the number of observations that are right censored at times after the jth event time, but prior to the $(j+1)$th time. From the log likelihood, $l(\pi_1, \ldots, \pi_J)$, the estimating equation for π_j is given by

$$\frac{\partial l(\pi_1, \ldots, \pi_J)}{\partial \pi_j} = \frac{d_j}{\pi_j} - \frac{n_j - d_j}{1 - \pi_j} = 0,$$

for $1 \leq j \leq J$. Then the maximum likelihood estimator of the jth conditional probability is obtained by solving the above estimating equation and is given by

$$\hat{\pi}_j = \frac{d_j}{n_j}.$$

Using the Central Limit Theorem, as n_j approaches infinity, the maximum likelihood estimator $\hat{\pi}_j$ has an asymptotic distribution given by

$$\sqrt{n_j}\left(\hat{\pi}_j - \pi_j\right) \xrightarrow{d} N\left(0, \pi_j(1 - \pi_j)\right).$$

Define time-conditional survival probability, $CS(b \mid a)$ for fixed times a and b such that $a < b$, as

$$CS(b \mid a) = \prod_{j: a < t_{(j)} \leq b} (1 - \pi_j).$$

Note that this has the same functional form that is used in the Kaplan-Meier estimator defined over

the range $a < t_{(j)} \leq b$. Like the survival function, the time-conditional survival estimator is a product of probabilities. We use the log transformation to obtain the mean and variance of the estimate. Define log time-conditional survival probability, for fixed times a and b such that $a < b$, as

$$\log CS(b \mid a) = \sum_{j:a<t_{(j)}\leq b} \log\left(1 - \pi_j\right),$$

as a sum of probabilities.

Applying the Taylor Series expansion to the log transformation given by $\log(1 - \pi_j)$, we have

$$\log(1 - \hat{\pi}_j) = \log(1 - \pi_j) - \frac{1}{1 - \pi}(\hat{\pi}_j - \pi_j) + R_2(a),$$

where

$$R_2(a) = (\hat{\pi}_j - \pi_j)^2 \cdot \frac{-1}{2(1 - \pi_j)^2}.$$

Asymptotically, as n_j approaches infinity, $R_2(a) \to 0$, since $\hat{\pi}_j$ is consistent for π_j. We derive the asymptotic distribution of this transformation by applying the Taylor Series expansion from the log transformation, $\log(1 - \pi_j)$, and, asymptotically, we have

$$\sqrt{n_j}\left(\log(1 - \hat{\pi}_j) - \log(1 - \pi_j)\right) = \sqrt{n_j} \cdot \frac{(\pi_j - \hat{\pi}_j)}{1 - \pi_j} + \sqrt{n_j} \cdot R_2(a), \tag{A.1}$$

where $R_2(a)$ is defined above. Since $\hat{\pi}_j$ follows an asymptotically normal distribution, the first term in Equation A.1 also follows an asymptotically normal distribution. Further, Lachin, 2000 shows that $\sqrt{n} \cdot R_2(a) \xrightarrow{p} 0$ by defining a sequence of values p_n as $n \to \infty$ and showing that p_n is a sample mean of n Bernoulli variables making it an $\sqrt{n}$-consistent estimator of π implying that $(p_n - \pi)^2 \to 0$ faster than $\sqrt{n}$. Using this information, we define $\sqrt{n_j}\left(\log(1 - \hat{\pi}_j) - \log(1 - \pi_j)\right)$ as the sum of two random variables shown above (Equation A.1). Of these, the first term converges in distribution to the normal distribution while the second term converges in probability to a constant of zero.

From Slutsky's Theorem for convergence in distribution, we have

$$\sqrt{n_j}\left(\log(1 - \hat{\pi}_j) - \log(1 - \pi_j)\right) \xrightarrow{d} N\left(0, \left(\frac{d\log(1 - \pi_j)}{d\pi_j}\right)^2 \cdot \pi_j(1 - \pi_j)\right),$$

where we have consistency given by

$$\log(1-\hat{\pi}_j) \xrightarrow{p} \log(1-\pi_j).$$

Then, by the δ-method, we have derived the first two large sample moments of this non-linear transformation as

$$E(\log(1-\hat{\pi}_j)) \cong \log(1-\pi_j),$$

and

$$Var(\log(1-\hat{\pi}_j)) \cong \frac{\pi_j}{n_j(1-\pi_j)}.$$

To derive the large sample asymptotic distribution for time-conditional survival probability, let n_j and n be large such that n_j/n converges in probability to ω_j, $n_j/n \xrightarrow{p} \omega_j$. Under this convergence in probability, the asymptotically normal distribution for the estimator of log time-conditional survival probability is given by

$$\sqrt{n}\frac{\left(\sum_{j:a<t_{(j)}\leq b}\log(1-\hat{\pi}_j) - \sum_{j:a<t_{(j)}\leq b}\log(1-\pi_j)\right)}{\sqrt{\sum_{j:a<t_{(j)}\leq b}\frac{\hat{\pi}_j}{\hat{\omega}_j(1-\hat{\pi}_j)}}} \xrightarrow{d} N(0,1).$$

Substituting the maximum likelihood estimator, $\hat{\pi}_j$, gives the estimated mean and variance as

$$\hat{E}\left(\log\widehat{CS}(b \mid a)\right) = \sum_{j:a<t_{(j)}\leq b}\log\left(1-\frac{d_j}{n_j}\right),$$

and

$$\widehat{Var}\left(\log\widehat{CS}(b \mid a)\right) = \sum_{j:a<t_{(j)}\leq b}\frac{d_j}{n_j(n_j-d_j)},$$

respectively. From the consistency of the MLE for π_j, a consistent estimator for the variance of time-conditional survival probability is given by

$$\widehat{Var}[\log\widehat{CS}(b \mid a)] \xrightarrow{p} \sum_{j:a<t_{(j)}\leq b}\frac{\pi_j}{\omega_j(1-\pi_j)}.$$

Peterson Jr, 1977 showed that, under random censoring, whatever the form of the true underlying

survival distribution, it is given that

$$\lim_{n\to\infty} \hat{S}(t) \xrightarrow{p} \lim_{n\to\infty} S(t).$$

Then, applying the logarithm transformation,

$$\lim_{n\to\infty} \log \hat{S}(t) \xrightarrow{p} \lim_{n\to\infty} \log S(t).$$

For two time points, a and b such that $b > a$ where $a, b \geq 0$, it holds that,

$$\lim_{n\to\infty} \log \hat{S}(b) - \log \hat{S}(a) \xrightarrow{p} \lim_{n\to\infty} \log S(b) - \log S(a),$$

and this is re-written to show

$$\lim_{n\to\infty} \widehat{CS}(b \mid a) \xrightarrow{p} \lim_{n\to\infty} CS(b \mid a).$$

This implies that, as the number of observations and, therefore, the number of events becomes larger, the intervals between steps in the nonparametric Kaplan-Meier step function estimating the probability become smaller and converge to the true underlying distribution. The successive conditional probabilities that estimate the conditional probability of surviving beyond an instant of time given survival up to that time, denoted as $\hat{\pi}_j$, are not statistically independent, but they are uncorrelated (Lachin, 2000). Further, this vector of probabilities is multivariate normally asymptotically distributed. In addition, the successive probabilities are asymptotically conditionally independent, that is, conditional on the numbers at risk at the preceding event times.

Define the general p-vector profile of log time-conditional survival probability estimators given by

$$\log \widehat{\mathbf{CS}} = \left(\log \widehat{CS}_1(b_1 \mid a_1), \log \widehat{CS}_2(b_2 \mid a_2), \ldots, \log \widehat{CS}_p(b_p \mid a_p) \right).$$

This profile is defined by a fixed difference between time b_i and time a_i where $i = 1, \ldots, p$ such that $b_i - a_i = c$, $\forall i$. Then the large sample asymptotic distribution for an estimated p-vector profile is given by

$$\log \widehat{\mathbf{CS}} \xrightarrow{d} N\left(E[\log \widehat{\mathbf{CS}}], Var[\log \widehat{\mathbf{CS}}] \right),$$

such that

$$E[\log \widehat{\mathbf{CS}}] \cong \log \mathbf{CS},$$

and

$$Var[\log \widehat{\mathbf{CS}}] \cong \boldsymbol{\Sigma},$$

where $\boldsymbol{\Sigma}$ is estimated as shown in Equations 2.10 and 2.11 and are derived below. The estimated mean and the estimated variance are above.

To determine the terms of the covariance matrix, Equations 2.8 and 2.9 define any two estimators of log time-conditional survival probabilities by

$$\log \widehat{CS}_l(b_l \mid a_l) \quad \text{and} \quad \log \widehat{CS}_m(b_m \mid a_m),$$

where $1 \leq l, m \leq J$ and where a_l, b_l, a_m, b_m are fixed times such that $0 \leq a_l \leq a_m \leq b_l \leq b_m \leq t_{(J)}$. The elements of the covariance matrix, $\boldsymbol{\Sigma}$, for $l = m$ are given by

$$\Sigma_{ll} = Var\left(\log \widehat{CS}_l(b_l \mid a_l)\right) = \sum_{j:a_l < t_{(j)} \leq b_l} \frac{\pi_j}{n_j(1-\pi_j)},$$

and for $l \neq m$ are given by

$$\Sigma_{lm} = Cov\left(\log \widehat{CS}_l(b_l \mid a_l), \log \widehat{CS}_m(b_m \mid a_m)\right) = \sum_{j:a_m < t_{(j)} \leq b_l} \frac{\pi_j}{n_j(1-\pi_j)}.$$

The covariance matrix, $\boldsymbol{\Sigma}$, is consistently estimated by $\hat{\boldsymbol{\Sigma}}$ where the elements of $\hat{\boldsymbol{\Sigma}}$ for $l = m$ are given by

$$\hat{\Sigma}_{ll} = \widehat{Var}\left(\log \widehat{CS}_l(b_l \mid a_l)\right) = \sum_{j:a_l < t_{(j)} \leq b_l} \frac{d_j}{n_j(n_j - d_j)}, \tag{A.2}$$

as shown earlier in Appendix A.2 and for $l \neq m$ are given by

$$\hat{\Sigma}_{lm} = \widehat{Cov}\left(\log \widehat{CS}_l(b_l \mid a_l), \log \widehat{CS}_m(b_m \mid a_m)\right) = \sum_{j:a_m < t_{(j)} \leq b_l} \frac{d_j}{n_j(n_j - d_j)}. \tag{A.3}$$

See Appendix Section A.2 for a detailed derivation. The covariance is 0 if a_l, b_l, a_m, b_m are non-overlapping fixed times such that $0 \leq a_l < b_l < a_m < b_m \leq t_{(J)}$.

APPENDIX B

CHAPTER 3

B.1. Partial derivatives for the Weibull distribution with continuous and categorical covariates

Below are partial derivatives of the parametric time-conditional survival probability for the Weibull distribution for a vector of covariates, $\mathbf{z} = (z_1, z_2, \ldots, z_p)^T$, where $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_p)^T$. Define time-conditional survival from this distribution by

$$CS = \frac{\exp\left(-b^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right)}{\exp\left(-a^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right)}$$

then the first degree partial derivatives are given by

$$\begin{aligned}
\frac{\partial CS}{\partial \alpha} &= \Big(\exp\left(-b^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \cdot \log(b) \cdot (-b^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})) \cdot \exp\left(-a^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \\
&\qquad - \exp\left(-a^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \cdot \log(a) \cdot (-a^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})) \cdot \exp\left(-b^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \Big) \\
&\quad \times \left(\exp\left(-2a^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right)\right)^{-1} \\
&= \Big(\exp\left(-b^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \cdot \log(b) \cdot (-b^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})) \\
&\qquad - \exp\left(-b^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \cdot \log(a) \cdot (-a^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})) \Big) \\
&\quad \times \left(\exp\left(-a^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right)\right)^{-1} \\
&= \frac{\exp\left(-b^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \lambda \exp(\boldsymbol{\beta}'\mathbf{z}) \left(a^\alpha \log(a) - b^\alpha \log(b)\right)}{\exp\left(-a^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right)},
\end{aligned}$$

$$\begin{aligned}
\frac{\partial CS}{\partial \lambda} &= \Big(-b^\alpha \exp(\boldsymbol{\beta}'\mathbf{z}) \exp\left(-b^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \exp\left(-a^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \\
&\qquad + a^\alpha \exp(\boldsymbol{\beta}'\mathbf{z}) \exp\left(-a^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \exp\left(-b^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \Big) \\
&\quad \times \left(\exp\left(-2a^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right)\right)^{-1} \\
&= \frac{a^\alpha \exp(\boldsymbol{\beta}'\mathbf{z}) \exp\left(-b^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) - b^\alpha \exp(\boldsymbol{\beta}'\mathbf{z}) \exp\left(-b^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right)}{\exp\left(-a^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right)} \\
&= \frac{\exp(\boldsymbol{\beta}'\mathbf{z}) \exp\left(-b^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \left(a^\alpha - b^\alpha\right)}{\exp\left(-a^\alpha \lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right)},
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial CS}{\partial \boldsymbol{\beta}} &= \Big((-b^{\alpha}\lambda \mathbf{z}) \exp\left(-b^{\alpha}\lambda \exp(\boldsymbol{\beta}'\mathbf{z}) + \boldsymbol{\beta}'\mathbf{z}\right) \exp\left(-a^{\alpha}\lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \\
&\qquad - (-a^{\alpha}\lambda \mathbf{z}) \exp\left(-a^{\alpha}\lambda \exp(\boldsymbol{\beta}'\mathbf{z}) + \boldsymbol{\beta}'\mathbf{z}\right) \exp\left(-b^{\alpha}\lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \Big) \\
&\quad \times \left(\exp\left(-2a^{\alpha}\lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right)\right)^{-1} \\
&= \frac{a^{\alpha}\lambda \mathbf{z} \exp(\boldsymbol{\beta}'\mathbf{z}) \exp\left(-b^{\alpha}\lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) - b^{\alpha}\lambda \mathbf{z} \exp(\boldsymbol{\beta}'\mathbf{z}) \exp\left(-b^{\alpha}\lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right)}{\exp\left(-a^{\alpha}\lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right)} \\
&= \frac{\lambda \mathbf{z} \exp(\boldsymbol{\beta}'\mathbf{z}) \exp\left(-b^{\alpha}\lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right) \left(a^{\alpha} - b^{\alpha}\right)}{\exp\left(-a^{\alpha}\lambda \exp(\boldsymbol{\beta}'\mathbf{z})\right)}.
\end{aligned}$$

These partial derivatives are used to derive the large variance of an estimate of time-conditional survival probability and to find the large sample covariance between any two distinct time-conditional survival probability estimates from the same sample.

B.2. Extending computations for the Logistic-Weibull with covariates

Below are the partial derivatives necessary to carry out the derivations described in Section 3.4 (based on Equation 3.13 and Equation 3.14). Define a time-conditional survival estimate from the Logistic-Weibull cure model extended to include multiple covariates as given by

$$\begin{aligned} &CS(b \mid a, \mathbf{x}, \mathbf{z}) \\ &\quad = \frac{\exp(\boldsymbol{\eta}'\mathbf{z}) \exp(-b^{\alpha} \cdot \lambda \cdot \exp(\boldsymbol{\zeta}'\mathbf{x})) + 1}{\exp(\boldsymbol{\eta}'\mathbf{z}) \exp(-a^{\alpha} \cdot \lambda \cdot \exp(\boldsymbol{\zeta}'\mathbf{x})) + 1} \end{aligned}$$

where $b = a + \Delta$, $\mathbf{z} = (1, z_1, z_2, \ldots, z_p)^T$, $\boldsymbol{\eta} = (\eta_0, \eta_1, \eta_2, \ldots, \eta_p)^T$, $\mathbf{x} = (x_1, x_2, \ldots, x_k)^T$, and $\boldsymbol{\zeta} = (\zeta_1, \zeta_2, \ldots, \zeta_k)^T$. Then the partial derivatives are given by

$$\begin{aligned} \frac{\partial CS}{\partial \boldsymbol{\eta}} = \Bigg(& \mathbf{z} \exp\left(-b^{\alpha} \lambda \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) \exp\left(\boldsymbol{\eta}'\mathbf{z}\right) \\ & \quad \times \left(\exp\left(\boldsymbol{\eta}'\mathbf{z}\right) \exp\left(-a^{\alpha} \lambda \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) + 1\right) \\ & - \mathbf{z} \exp\left(-a^{\alpha} \lambda \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) \exp\left(\boldsymbol{\eta}'\mathbf{z}\right) \\ & \quad \times \left(\exp\left(\boldsymbol{\eta}'\mathbf{z}\right) \exp\left(-b^{\alpha} \lambda \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) + 1\right) \Bigg) \\ & \times \left(\exp\left(\boldsymbol{\eta}'\mathbf{z}\right) \exp\left(-a^{\alpha} \lambda \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) + 1\right)^{-2}, \end{aligned}$$

$$\begin{aligned} \frac{\partial CS}{\partial \lambda} = \Bigg(& - \exp\left(\boldsymbol{\eta}'\mathbf{z}\right) \cdot b^{\alpha} \cdot \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right) \\ & \quad \times \exp\left(-b^{\alpha} \lambda \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) \\ & \quad \times \left(\exp\left(\boldsymbol{\eta}'\mathbf{z}\right) \exp\left(-a^{\alpha} \lambda \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) + 1\right) \\ & + \exp\left(\boldsymbol{\eta}'\mathbf{z}\right) \cdot a^{\alpha} \cdot \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right) \\ & \quad \times \exp\left(-a^{\alpha} \lambda \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) \\ & \quad \times \left(\exp\left(\boldsymbol{\eta}'\mathbf{z}\right) \exp\left(-b^{\alpha} \lambda \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) + 1\right) \Bigg) \\ & \times \left(\exp\left(\boldsymbol{\eta}'\mathbf{z}\right) \exp\left(-a^{\alpha} \lambda \exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) + 1\right)^{-2}, \end{aligned}$$

$$\begin{aligned}\frac{\partial CS}{\partial \alpha} = \Bigg(& -\exp\left(\boldsymbol{\eta}'\mathbf{z}\right)\lambda\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right) \\ & \quad \times b^{\alpha}\cdot\log b\exp\left(-b^{\alpha}\lambda\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) \\ & \quad \times\left(\exp\left(\boldsymbol{\eta}'\mathbf{z}\right)\exp\left(-a^{\alpha}\lambda\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right)+1\right) \\ & +\exp\left(\boldsymbol{\eta}'\mathbf{z}\right)\lambda\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right) \\ & \quad \times a^{\alpha}\cdot\log a\exp\left(-a^{\alpha}\lambda\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) \\ & \quad \times\left(\exp\left(\boldsymbol{\eta}'\mathbf{z}\right)\exp\left(-b^{\alpha}\lambda\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right)+1\right)\Bigg) \\ & \times\left(\exp\left(\boldsymbol{\eta}'\mathbf{z}\right)\exp\left(-a^{\alpha}\lambda\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right)+1\right)^{-2},\end{aligned}$$

and

$$\begin{aligned}\frac{\partial CS}{\partial \boldsymbol{\zeta}} = \Bigg(& -\exp\left(\boldsymbol{\eta}'\mathbf{z}\right)\cdot b^{\alpha}\cdot\lambda\cdot\mathbf{x}\cdot\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right) \\ & \quad \times\exp\left(-b^{\alpha}\lambda\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) \\ & \quad \times\left(\exp\left(\boldsymbol{\eta}'\mathbf{z}\right)\exp\left(-a^{\alpha}\lambda\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right)+1\right) \\ & +\exp\left(\boldsymbol{\eta}'\mathbf{z}\right)\cdot a^{\alpha}\cdot\lambda\cdot\mathbf{x}\cdot\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right) \\ & \quad \times\exp\left(-a^{\alpha}\lambda\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right) \\ & \quad \times\left(\exp\left(\boldsymbol{\eta}'\mathbf{z}\right)\exp\left(-b^{\alpha}\lambda\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right)+1\right)\Bigg) \\ & \times\left(\exp\left(\boldsymbol{\eta}'\mathbf{z}\right)\exp\left(-a^{\alpha}\lambda\exp\left(\boldsymbol{\zeta}'\mathbf{x}\right)\right)+1\right)^{-2}.\end{aligned}$$

B.3. Cure model sensitivity analyses

The objective of the sensitivity analyses was to assess the effect of including a covariate in the survival and/or the probability components of the cure model on the time-conditional survival probability estimates. While we assessed several models, the results from four models are presented here to illustrate the relationship.

- M1: Thickness (Logistic and Weibull)
- M2: Thickness (Logistic and Weibull) and Ulceration (Weibull)
- M3: Thickness (Logistic and Weibull) and Ulceration (Logistic)
- M4: Thickness (Logistic and Weibull) and Ulceration (Logistic and Weibull)

We first start by looking at the estimated β values from the cure model (Table B.1). In looking at the Logistic component of the cure model, the range of estimates for thickness was from 0.1264 (M1) up to 0.1337 (M4). For the Weibull component of the cure model, the range of estimates for thickness was from 0.0771 (M2) up to 0.0788 (M3). After including ulceration in the Weibull component only (M2), the estimate of thickness in the Weibull component was the smallest among the 4 models and the estimate of thickness in the Logistic component increased slightly from M1. Including ulceration in the Logistic component only (M3), the estimate of thickness in the Weibull component was the largest among the 4 models and the estimate of thickness in the Logistic component was greater than in M1 and in M2.

After including ulceration in both components of the model (M4), the estimate of thickness in the Weibull component was greater than in M1 and M2 and the estimate of thickness in the Logistic component was the largest among the four models. Thickness remained significant in all four models (all $p < .0001$) irrespective of the inclusion of ulceration. Ulceration was significant in all four models though the p-value was slightly greater (closer to the critical value of 0.05) when included in both components in M4 ($p < .0001$ in the Logistic component in M3 versus $p = 0.0018$ in M4 and $p < .0001$ in the Weibull component in M2 versus $p = 0.0007$ in the Weibull component in M4).

Assessing the Akaike information criterion (AIC) and Bayesian information criterion (BIC) measures of the four models, M4 (ulceration in both components) had the smallest AIC at 8751.7 and

the smallest BIC at 8794.8. The largest estimated AIC and BIC were for M1 (thickness in both components) at 8774.7 and 8805.4, respectively. M2 had the second smallest estimated AIC and BIC at 8759.4 and 8796.3, respectively, and M3 has the second largest estimated AIC and BIC at 8760.6 and 8797.6, respectively.

As these models are nest, we computed the likelihood ratio test given by

$$G = -2 * (logL(reducedmodel)) + 2 * (logL(fullmodel)).$$

Table B.2 shows the results of the calculations. When the model with ulceration in both components (M4) was compared to the model with ulceration in the Logistic component of the cure model only (M3), the test statistic was 10.9 and the p-value was 0.0010. Similarly, when the model with ulceration in both components (M4) was compared to the model with ulceration in the Weibull component of the cure model only (M2), the test statistic was 9.7 and the p-value was 0.0018. This indicated that the model with ulceration in both components is better than each of the models with ulceration in just one component.

Further, we compared the model without ulceration (M1) to the model with ulceration in the Logistic component of the cure model only (M3) and found the test statistic was 16.1 and the p-value was ¡.0001. Similarly, when the model without ulceration (M1) was compared to the model with ulceration in the Weibull component of the cure model only (M2), the test statistic was 17.3 and the p-value was ¡.0001. This indicated that the model without ulceration has poorer fit to these data than each of the models with ulceration in just one component.

Next, we assessed the effect of ulceration across these four models on the estimates of time-conditional survival. As in the analysis presented in the chapter, we fixed tumor thickness at 3.58 mm and varied ulceration status (with and without ulceration). We present the estimated 5-year time-conditional survival probabilities given survival beyond 1 and beyond 10 years after diagnosis in Table B.3. For these data, we found the pairwise comparisons of 5-year time conditional survival given survival beyond 1 year versus given survival beyond 10 years after diagnosis were significantly different in all the models evaluated.

From the 7 comparisons, the estimate of 5-year time-conditional survival given survival beyond 1 year after diagnosis and also the smallest estimate given survival beyond 10 years after diagnosis was smallest for comparison 5 under M3 (ulceration in the Logistic component of the cure model only) with ulceration present (0.9068 and 0.9935, respectively). The estimated difference in 5-year time-conditional survival given 1 and given 10 years after diagnosis was the largest among the 7 comparisons at 0.0867 and the hypothesis test indicated there was a significant difference in the two time-conditional survival estimates ($p < .0001$). The largest estimate of 5-year time-conditional survival given 1 year after diagnosis and also the largest estimate given 10 years after diagnosis was for comparison 3 under M2 (ulceration in the Weibull component of the cure model only) with ulceration present (0.9389 and 0.9966, respectively). The estimated difference in 5-year time-conditional survival given 1 and given 10 years after diagnosis was the smallest for this comparison among all 7 comparisons at 0.0577 and the hypothesis test indicated there was a significant difference in the two time-conditional survival estimates ($p < .0001$).

While the smallest observed difference in 5-year time-conditional survival estimates was observed for comparison 3 for M2 with ulceration present (ulceration in the Weibull component), this was closest in magnitude to the three comparisons for patients without ulceration ranging at 0.0636 for comparison 4 (M3: ulceration in the Logistic component only), 0.0641 for comparison 6 (M4: ulceration in both components), and 0.0693 for comparison 2 (M2: ulceration in the Weibull component only). The larger estimated differences in 5-year time-conditional survival given survival beyond 1 and 10 years after diagnosis were 0.0701 for comparison 1 (M1: thickness in both components), 0.0707 for comparison 7 (M4: ulceration in both components), and 0.0867 for comparison 5 (M3: ulceration in the Logistic component only).

Overall, for M2 with ulceration in the Weibull component only, patients with a tumor thickness of 3.58 mm (the sample average) and with ulceration saw a smaller increase in estimated time-conditional survival relative to those without ulceration (0.0577 for comparison 3 versus 0.0693 for comparison 2). All models for patients with a tumor thickness of 3.58 mm and without ulceration saw a smaller difference in 5-year time-conditional survival relative to the model with tumor thickness only (comparisons 4, 6, and 2 had smaller estimates relative to comparison 1). Further, for M3 with ulceration in the Logistic component only, patients with a tumor thickness of 3.58 mm and without ulceration had a higher initial estimated 5-year time-conditional survival probability given 1 year after survival

and saw a smaller increase in estimated difference in time-conditional survival relative to those with ulceration (0.0636 for comparison 4 versus 0.0867 for comparison 5). For M4 with ulceration in both components, patients with a tumor thickness of 3.58mm and without ulceration also had a higher initial estimated 5-year time-conditional survival probability given 1 year after survival and saw a smaller increase in estimated difference in time-conditional survival relative to those with ulceration (0.0641 for comparison 6 versus 0.0707 for comparison 7).

Table B.1: Maximum likelihood estimates from four Weibull mixture cure models for disease-specific survival.

Variable	M1	M2	M3	M4
Logistic Model				
Intercept	-1.0310^{*}	-1.0373^{*}	-1.1555^{*}	-1.142^{*}
Thickness (mm)	0.1264^{*}	0.1289^{*}	0.1323^{*}	0.1337^{*}
Ulceration (vs without)	–	–	0.3472^{*}	0.2743^{**}
Weibull Survival Model				
Intercept (Weibull)	1.8716^{*}	1.9446^{*}	1.8787^{*}	1.9301^{*}
Thickness (mm)	0.0773^{*}	0.0771^{*}	0.0788^{*}	0.0774^{*}
Ulceration (vs without)	–	0.3293^{*}	–	0.2753^{***}
Shape (Weibull)	0.6915^{*}	0.6882^{*}	0.6920^{*}	0.6870^{*}

*p-value $< .0001$, **p-value $= 0.0018$, *** p-value $= 0.0007$

Table B.2: Results from the likelihood ratio test for nested models.

Reduced Model	Full Model	Test Statistic	p-value
M3	M4	10.9	0.0010
M2	M4	9.7	0.0018
M1	M3	16.1	$< .0001$
M1	M2	17.3	$< .0001$

Table B.3: Estimates of 5-year time-conditional survival probability from four Weibull mixture cure models for disease-specific survival adjusting for fixed tumor thickness (3.58mm) and varying ulceration status.

Comparison i	Ulceration Status	$\widehat{CS}(6 \mid 1)$ $(j = 1)$	$\widehat{CS}(15 \mid 10)$ $(j = 2)$	Estimates (H_1)	p-value
1 (M1)	–	0.9248	0.9949	0.0701	$< .0001$
2 (M2)	0	0.9248	0.9941	0.0693	$< .0001$
3 (M2)	1	0.9389	0.9966	0.0577	$< .0001$
4 (M3)	0	0.9318	0.9954	0.0636	$< .0001$
5 (M3)	1	0.9068	0.9935	0.0867	$< .0001$
6 (M4)	0	0.9306	0.9947	0.0641	$< .0001$
7 (M4)	1	0.9249	0.9956	0.0707	$< .0001$

APPENDIX C

CHAPTER 4

C.1. Partial derivatives

Recall that our assumed likelihood is given by,

$$L(\beta,\alpha) \quad = \quad \prod_{i=1}^{m} \exp\left(y_{i1}\ln(\lambda_{i1}) - \lambda_{i1} - \ln(y_{i1}!)\right) \prod_{j=2}^{n_i} \exp\left(y_{ij}\ln({\lambda_{ij}}^*) - {\lambda_{ij}}^* - \ln(y_{ij}!)\right).$$

Define $\ell(\beta,\alpha) = Ln(L(\beta,\alpha))$. Taking natural logs then yields,

$$\begin{aligned} \ell(\beta,\alpha) &= Ln(L(\beta,\alpha)) \\ &= \sum_{i=1}^{m} \left(y_{i1}\theta_{i1} - \exp(\theta_{i1}) - \ln(y_{i1}!)\right) \\ &\quad + \left(y_{i2}{\theta_{i2}}^* - \exp({\theta_{i2}}^*) - \ln(y_{i2}!)\right) \\ &\quad + \sum_{j=3}^{n_i} \left(y_{ij}{\theta_{ij}}^* - \exp({\theta_{ij}}^*) - \ln(y_{ij}!)\right), \end{aligned} \tag{C.1}$$

where $\theta_{i1} = \ln(\lambda_{i1}) = x'_{i1}\beta$ and ${\theta_{ij}}^* = \ln({\lambda_{ij}}^*)$ for $j = 2,\ldots,n_i$.

As noted earlier, to obtain maximum likelihood estimates of β and of α we need to solve the following estimating equations:

$$\frac{\partial \ln\left(L(\beta,\alpha)\right)}{\partial\beta} \quad = \quad 0$$

and

$$\frac{\partial \ln\left(L(\beta,\alpha)\right)}{\partial\alpha} \quad = \quad 0.$$

The partial derivative for the estimating equations with respect to β is given by,

$$\begin{aligned}\frac{\partial}{\partial\beta}\ell(\beta,\alpha) = \sum_{i=1}^{m} & y_{i1}x_{i1} - \lambda_{i1}x_{i1} \\ & + \left(x_{i2}\lambda_{i2} + \frac{\alpha}{\sqrt{1-\alpha^2}}\frac{\sqrt{\lambda_{i2}}}{2}\left(\frac{y_{i1}}{\sqrt{\lambda_{i1}}}(x_{i2}-x_{i1}) - \sqrt{\lambda_{i1}}(x_{i2}+x_{i1})\right)\right) \\ & \quad \times \left(y_{i2}\left(\lambda_{i2} + \frac{\alpha}{\sqrt{1-\alpha^2}}\sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(y_{i1}-\lambda_{i1})\right)^{-1} - 1\right) \\ & + \sum_{j=3}^{n_i}\left(x_{ij}\lambda_{ij} + \frac{\alpha\sqrt{\lambda_{ij}}}{2}\left(\frac{y_{ij-1}}{\sqrt{\lambda_{ij-1}}}(x_{ij}-x_{ij-1}) - \sqrt{\lambda_{ij-1}}(x_{ij}+x_{ij-1})\right)\right) \\ & \quad \times \left(y_{ij}\left(\lambda_{ij} + \alpha\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}(y_{ij-1}-\lambda_{ij-1})\right)^{-1} - 1\right),\end{aligned} \tag{C.2}$$

and the partial derivative with respect to α is given by

$$\begin{aligned}\frac{\partial}{\partial\alpha}\ell(\beta,\alpha) = \sum_{i=1}^{m} & \left((1-\alpha^2)^{-3/2}\sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(y_{i1}-\lambda_{i1})\right) \\ & \quad \times \left(y_{i2}\left(\lambda_{i2} + \frac{\alpha}{\sqrt{1-\alpha^2}}\sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(y_{i1}-\lambda_{i1})\right)^{-1} - 1\right) \\ & + \sum_{j=3}^{n_i}\left(\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}(y_{ij-1}-\lambda_{ij-1})\right) \\ & \quad \times \left(y_{ij}\left(\lambda_{ij} + \alpha\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}(y_{ij-1}-\lambda_{ij-1})\right)^{-1} - 1\right).\end{aligned} \tag{C.3}$$

The elements of the matrix of second-order partial derivatives of the log likelihood, called the Hes-

sian matrix, are given by,

$$
\begin{aligned}
\frac{\partial^2 \ell(\beta,\alpha)}{\partial\beta\partial\beta'} &= \sum_{i=1}^{m}(-x_{i1}x_{i1}\lambda_{i1}) \\
&+ \left(x_{i2}x_{i2}\lambda_{i2} + \left(\frac{\alpha}{\sqrt{1-\alpha^2}}\frac{\sqrt{\lambda_{i2}}}{2}\frac{x_{i2}}{2}\right) \times \left(\frac{y_{i1}(x_{i2}-x_{i1})}{\sqrt{\lambda_{i1}}} - (x_{i2}+x_{i1})\sqrt{\lambda_{i1}}\right)\right. \\
&\quad \left. - \left(\frac{y_{i1}(x_{i2}-x_{i1})x_{i1}}{2\sqrt{\lambda_{i1}}} + \frac{(x_{i2}+x_{i1})x_{i1}\sqrt{\lambda_{i1}}}{2}\right) \times \left(\frac{\alpha}{\sqrt{1-\alpha^2}}\frac{\sqrt{\lambda_{i2}}}{2}\right)\right) \\
&\quad \times \left(y_{i2}\left(\lambda_{i2} + \frac{\alpha}{\sqrt{1-\alpha^2}}\sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(y_{i1}-\lambda_{i1})\right)^{-1} - 1\right)\Bigg) \\
&- \left(y_{i2}\left(\lambda_{i2} + \frac{\alpha}{\sqrt{1-\alpha^2}}\sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(y_{i1}-\lambda_{i1})\right)^{-2}\right. \\
&\quad \left. \times \left(x_{i2}\lambda_{i2} + \frac{\alpha}{\sqrt{1-\alpha^2}}\left(\frac{1}{2}\sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(x_{i2}-x_{i1})(y_{i1}-\lambda_{i1}) - x_{i1}\lambda_{i1}\sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}\right)\right)\right) \\
&\quad \times \left(\left(x_{i2}\lambda_{i2} + \frac{\alpha}{\sqrt{1-\alpha^2}}\frac{\sqrt{\lambda_{i2}}}{2}\left(\frac{y_{i1}}{\sqrt{\lambda_{i1}}}(x_{i2}-x_{i1}) - \sqrt{\lambda_{i1}}(x_{i2}+x_{i1})\right)\right)\right) \qquad \text{(C.4)} \\
&+ \sum_{j=3}^{n_i} x_{ij}x_{ij}\lambda_{ij} + \left(\alpha\frac{\sqrt{\lambda_{ij}}}{2}\frac{x_{ij}}{2}\right) \times \left(\frac{y_{ij-1}(x_{ij}-x_{ij-1})}{\sqrt{\lambda_{ij-1}}} - (x_{ij}+x_{ij-1})\sqrt{\lambda_{ij-1}}\right) \\
&- \left(\frac{y_{ij-1}(x_{ij}-x_{ij-1})x_{ij-1}}{2\sqrt{\lambda_{ij-1}}} + \frac{(x_{ij}+x_{ij-1})x_{ij-1}\sqrt{\lambda_{ij-1}}}{2}\right) \times \left(\alpha\frac{\sqrt{\lambda_{ij}}}{2}\right) \\
&\quad \times \left(y_{ij}\left(\lambda_{ij} + \alpha\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}(y_{ij-1}-\lambda_{ij-1})\right)^{-1} - 1\right)\Bigg) \\
&- \left(y_{ij}\left(\lambda_{ij} + \alpha\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}(y_{ij-1}-\lambda_{ij-1})\right)^{-2}\right. \\
&\quad \left. \times \left(x_{ij}\lambda_{ij} + \alpha\left(\frac{1}{2}\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}(x_{ij}-x_{ij-1})(y_{ij-1}-\lambda_{ij-1}) - x_{ij-1}\lambda_{ij-1}\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}\right)\right)\right) \\
&\quad \times \left(x_{ij}\lambda_{ij} + \alpha\frac{\sqrt{\lambda_{ij}}}{2}\left(\frac{y_{ij-1}}{\sqrt{\lambda_{ij-1}}}(x_{ij}-x_{ij-1}) - \sqrt{\lambda_{ij-1}}(x_{ij}+x_{ij-1})\right)\right),
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \ell(\beta,\alpha)}{\partial\beta\partial\alpha} = \sum_{i=1}^{m} & \left((1-\alpha^2)^{-3/2} \left(\frac{1}{2}\sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(x_{i2}-x_{i1})(y_{i1}-\lambda_{i1}) - x_{i1}\lambda_{i1}\sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}} \right) \right) \\
& \times \left(y_{i2} \left(\lambda_{i2} + \frac{\alpha}{\sqrt{1-\alpha^2}} \sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(y_{i1}-\lambda_{i1}) \right)^{-1} - 1 \right) \\
& - \left(y_{i2} \left(\lambda_{i2} + \frac{\alpha}{\sqrt{1-\alpha^2}} \sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(y_{i1}-\lambda_{i1}) \right)^{-2} \right. \\
& \qquad \left. \times \left(x_{i2}\lambda_{i2} + \frac{\alpha}{\sqrt{1-\alpha^2}} \left(\frac{1}{2}\sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(x_{i2}-x_{i1})(y_{i1}-\lambda_{i1}) - x_{i1}\lambda_{i1}\sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}} \right) \right) \right) \\
& \times \left((1-\alpha^2)^{-3/2} \sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(y_{i1}-\lambda_{i1}) \right) \\
& + \sum_{j=3}^{n_i} \left(\left(\frac{1}{2}\sqrt{\frac{\lambda_{i2j}}{\lambda_{ij-1}}}(x_{ij}-x_{ij-1})(y_{ij-1}-\lambda_{ij-1}) - x_{ij-1}\lambda_{ij-1}\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}} \right) \right) \\
& \times \left(y_{ij} \left(\lambda_{ij} + \alpha\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}(y_{ij-1}-\lambda_{ij-1}) \right)^{-1} - 1 \right) \\
& - \left(y_{ij} \left(\lambda_{ij} + \alpha\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}(y_{ij-1}-\lambda_{ij-1}) \right)^{-2} \right. \\
& \qquad \left. \times \left(x_{ij}\lambda_{ij} + \alpha \left(\frac{1}{2}\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}(x_{ij}-x_{ij-1})(y_{ij-1}-\lambda_{ij-1}) - x_{ij-1}\lambda_{ij-1}\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}} \right) \right) \right) \\
& \times \left(\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}(y_{ij-1}-\lambda_{ij-1}) \right),
\end{aligned}
\tag{C.5}
$$

and

$$
\begin{aligned}
\frac{\partial^2 \ell(\beta,\alpha)}{\partial\alpha^2} = \sum_{i=1}^{m} & \left(\sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(y_{i1}-\lambda_{i1}) \frac{3\alpha}{(1-\alpha^2)^{5/2}} \right) \\
& \times \left(y_{i2} \left(\lambda_{i2} + \frac{\alpha}{\sqrt{1-\alpha^2}} \sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(y_{i1}-\lambda_{i1}) \right)^{-1} - 1 \right) \\
& - y_{i2} \left(\lambda_{i2} + \frac{\alpha}{\sqrt{1-\alpha^2}} \sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(y_{i1}-\lambda_{i1}) \right)^{-2} \\
& \qquad \times \left(\sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(y_{i1}-\lambda_{i1})(1-\alpha^2)^{-3/2} \right) \\
& \times \left((1-\alpha^2)^{-3/2} \sqrt{\frac{\lambda_{i2}}{\lambda_{i1}}}(y_{i1}-\lambda_{i1}) \right) \\
& + \sum_{j=3}^{n_i} \left(- y_{ij} \left(\lambda_{ij} + \alpha\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}(y_{ij-1}-\lambda_{ij-1}) \right)^{-2} \right. \\
& \qquad \left. \times \left(\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}(y_{ij-1}-\lambda_{ij-1}) \right) \right) \\
& \times \left(\sqrt{\frac{\lambda_{ij}}{\lambda_{ij-1}}}(y_{ij-1}-\lambda_{ij-1}) \right).
\end{aligned}
\tag{C.6}
$$

C.2. R function for log likelihood

```
#####################################################################
## Log Likelihood function
## This function was written by Victoria Gamerman
#####################################################################

logl3 <- function(start.values){
 alpha <- start.values[1]
 beta <- start.values[2:length(start.values)]
 formula <- y ~ trt + base + age
# id <- epil$id
# time <- epil$period
 d <- dim(epil)
 k <- length(all.vars(formula))-1
 dt.fm<- data.frame(epil)

dataset<- data.proc(data=dt.fm,formula=formula,time=time,id=id,del.n=0)
  m<- dataset$m
  n<- dataset$n
  id<- dataset$id
  time<- dataset$time

l_beta_a <- 0
l_beta_b <- 0
l_beta_c <- 0
for (i in 1:m){
        data_i <- matrix(NA, nrow=n[i], ncol=dim(dataset$data)[2])
        data_i[1:n[i],1:dim(dataset$data)[2]] <- dataset$data[which(id==i),]
        data.end<- ncol(data_i)
        x_i <- matrix(NA, nrow=n[i], ncol=k+1)
        x_i[1:n[i],1:(k+1)] <- data_i[,-data.end]
        y_i<- data_i[,data.end]
        n_i <- nrow(data_i)

for (j in 1:n_i){
  if (j == 1){
    lam_ij <- exp(t(beta)%*%x_i[j,])
    lam_ij <- lam_ij[1]
    l_beta_a <- l_beta_a + y_i[j]*log(lam_ij) - exp(log(lam_ij)) - log(factorial(y_i[j]))
  }
  if (j == 2){
    lam_ij <- exp(t(beta)%*%x_i[j,])
    lam_ij <- lam_ij[1]
```

```
    lam_ij_1 <- exp(t(beta)%*%x_i[j-1,])
    lam_ij_1 <- lam_ij_1[1]
    lamdot_i2 <- lam_ij + (alpha / sqrt(1-alpha^2))*sqrt(lam_ij / lam_ij_1)*(y_i[j-1] - lam_ij_
        1)
      if(is.finite(lamdot_i2) == FALSE){ lamdot_i2 <- 0.2} #constraint check
      if(lamdot_i2 < 0){lamdot_i2 <- 0.2} #constraint check
    l_beta_b <- l_beta_b + y_i[j]*log(lamdot_i2) - exp(log(lamdot_i2)) - log(factorial(y_i[j]))
  }
  if (j > 2){
    lam_ij <- exp(t(beta)%*%x_i[j,])
    lam_ij <- lam_ij[1]
    lam_ij_1 <- exp(t(beta)%*%x_i[j-1,])
    lam_ij_1 <- lam_ij_1[1]
    lamdot_ij <- lam_ij + alpha *sqrt(lam_ij / lam_ij_1)*(y_i[j-1] - lam_ij_1)
      if(is.finite(lamdot_ij) == FALSE){ lamdot_ij <- 0.2} #constraint check
      if(lamdot_ij < 0){ lamdot_ij <- 0.2} #constraint check
    l_beta_c <- l_beta_c + y_i[j]*log(lamdot_ij) - exp(log(lamdot_ij)) - log(factorial(y_i[j]))
  }
}
}
loglik <- l_beta_a + l_beta_b + l_beta_c
return(loglik)
}
```

C.3. R function for gradient

```
#####################################################################
## Gradient function: It should take arguments matching those of f
## and return a vector containing the gradient.
## This function was written by Victoria Gamerman
#####################################################################

ml.grad <- function(start.values){
 alpha <- start.values[1]
 beta <- start.values[2:length(start.values)]
data<-epil
# formula <- y_sim ~ trt + base + age
# time <- data$time
# id <- data$id
d <- dim(data)
k <- length(all.vars(formula))-1
dt.fm<- data.frame(data)
dataset<- data.proc(data=dt.fm,formula=formula,time=time,id=id,del.n=0)
  m<- dataset$m
  n<- dataset$n
  id<- dataset$id
  time<- dataset$time
  autotime<- dataset$autotime

l_beta_a <- matrix(0,nrow=k+1, ncol=1)
l_beta_b <- matrix(0,nrow=k+1, ncol=1)
l_beta_c <- matrix(0,nrow=k+1, ncol=1)
l_alpha_a <- matrix(0,nrow=1, ncol=1)
l_alpha_b <- matrix(0,nrow=1, ncol=1)

for (i in 1:m){
  data_i <- matrix(NA, nrow=n[i], ncol=dim(dataset$data)[2])
  data_i[1:n[i],1:dim(dataset$data)[2]] <- dataset$data[which(id==i),]
  data.end<- ncol(data_i)
  x_i <- matrix(NA, nrow=n[i], ncol=k+1)
  x_i[1:n[i],1:(k+1)] <- data_i[,-data.end]
  y_i<- data_i[,data.end]
  n_i <- nrow(data_i)

if (n_i>=1){
  for (j in 1:n_i){
    if (j == 1){
      lam_ij <- exp(t(beta)%*%x_i[j,])
      lam_ij <- lam_ij[1]
      l_beta_a <- l_beta_a + y_i[j]*x_i[j,]-x_i[j,]*lam_ij
```

```
    }
    if(j==2){
      lam_ij <- exp(t(beta)%*%x_i[j,])
      lam_ij <- lam_ij[1]
      lam_ij_1 <- exp(t(beta)%*%x_i[j-1,])
      lam_ij_1 <- lam_ij_1[1]
      l_alpha_a <- l_alpha_a + y_i[j]*(lam_ij + (alpha/sqrt(1-alpha^2))*sqrt(lam_ij/lam_ij_1)*(
          y_i[j-1]-lam_ij_1))^(-1)*
        (sqrt(lam_ij/lam_ij_1)*(y_i[j-1]-lam_ij_1)*((1-alpha^2)^(-3/2)))-(sqrt(lam_ij/lam_ij_1)
            *(y_i[j-1]-lam_ij_1)*((1-alpha^2)^(-3/2)))
      l_beta_b <- l_beta_b + y_i[j]*(lam_ij+(alpha/sqrt(1-alpha^2))*sqrt(lam_ij/lam_ij_1)*(y_i[
          j-1]-lam_ij_1))^(-1)*
        (x_i[j,]*lam_ij+(alpha/sqrt(1-alpha^2))*(0.5*sqrt(lam_ij/lam_ij_1)*(x_i[j,]-x_i[j-1,])*
            (y_i[j-1]-lam_ij_1)-x_i[j-1,]*lam_ij_1*sqrt(lam_ij/lam_ij_1)))-
      (x_i[j,]*lam_ij+(alpha/sqrt(1-alpha^2))*(0.5*sqrt(lam_ij/lam_ij_1)*(x_i[j,]-x_i[j-1,])*(y
          _i[j-1]-lam_ij_1)-x_i[j-1,]*lam_ij_1*sqrt(lam_ij/lam_ij_1)))
    }
    if(j>2){
      lam_ij <- exp(t(beta)%*%x_i[j,])
      lam_ij <- lam_ij[1]
      lam_ij_1 <- exp(t(beta)%*%x_i[j-1,])
      lam_ij_1 <- lam_ij_1[1]
      l_alpha_b <- l_alpha_b + y_i[j]*(lam_ij + alpha*sqrt(lam_ij/lam_ij_1)*(y_i[j-1]-lam_ij_1)
          )^(-1)*
        (sqrt(lam_ij/lam_ij_1)*(y_i[j-1]-lam_ij_1))-(sqrt(lam_ij/lam_ij_1)*(y_i[j-1]-lam_ij_1))
      l_beta_c <- l_beta_c + y_i[j]*(lam_ij + alpha*sqrt(lam_ij/lam_ij_1)*(y_i[j-1]-lam_ij_1))
          ^(-1)*
        (x_i[j,]*lam_ij+alpha*(0.5*sqrt(lam_ij/lam_ij_1)*(x_i[j,]-x_i[j-1,])*(y_i[j-1]-lam_ij_
            1)-x_i[j-1,]*lam_ij_1*sqrt(lam_ij/lam_ij_1)))-
      (x_i[j,]*lam_ij+alpha*(0.5*sqrt(lam_ij/lam_ij_1)*(x_i[j,]-x_i[j-1,])*(y_i[j-1]-lam_ij_1)-
          x_i[j-1,]*lam_ij_1*sqrt(lam_ij/lam_ij_1)))
    }
  }
}
}
l_alpha <- l_alpha_a+l_alpha_b
l_beta <- l_beta_a+l_beta_b+l_beta_c
out<-t(t(c(l_alpha,l_beta)))

return(out)
}
```

C.4. R function for score

```
################################################################################
## Score Squared function
## This function was written by Victoria Gamerman
################################################################################

score_squared <- function(formula, data, id, time, alpha, beta){
  formula <- y ~ trt + base + age
  data<-epil
  id<-epil$subject
  time<-epil$period
  alpha<-mle.alpha
  beta<-mle.beta

d <- dim(data)
k <- length(all.vars(formula))-1
dt.fm<- data.frame(data)
dataset<- data.proc(data=dt.fm,formula=formula,time=time,id=id,del.n=0)
  m<- dataset$m
  n<- dataset$n
  id<- dataset$id
  time<- dataset$time
  autotime<- dataset$autotime
p <- length(beta) + 1
squared <- matrix(0, nrow=p, ncol=p)

out.score <- matrix(NA, nrow=m, ncol=p+1)

for (i in 1:m){
  l_beta_a <- matrix(0,nrow=k+1, ncol=1)
  l_beta_b <- matrix(0,nrow=k+1, ncol=1)
  l_beta_c <- matrix(0,nrow=k+1, ncol=1)
  l_alpha_a <- matrix(0,nrow=1, ncol=1)
  l_alpha_b <- matrix(0,nrow=1, ncol=1)
  data_i <- matrix(NA, nrow=n[i], ncol=dim(dataset$data)[2])
  data_i[1:n[i],1:dim(dataset$data)[2]] <- dataset$data[which(id==i),]
  data.end<- ncol(data_i)
  x_i <- matrix(NA, nrow=n[i], ncol=k+1)
  x_i[1:n[i],1:(k+1)] <- data_i[,-data.end]
  y_i<- data_i[,data.end]
  n_i <- nrow(data_i)
if (n_i>=1){
  for (j in 1:n_i){
    if (j == 1){
      lam_ij <- exp(t(beta)%*%x_i[j,])
```

```
45        lam_ij <- lam_ij[1]
46        l_beta_a <- l_beta_a + y_i[j]*x_i[j,]-x_i[j,]*lam_ij
47      }
48      if(j==2){
49        lam_ij <- exp(t(beta)%*%x_i[j,])
50        lam_ij <- lam_ij[1]
51        lam_ij_1 <- exp(t(beta)%*%x_i[j-1,])
52        lam_ij_1 <- lam_ij_1[1]
53        l_alpha_a <- l_alpha_a + y_i[j]*(lam_ij + (alpha/sqrt(1-alpha^2))*sqrt(lam_ij/lam_ij_1)*(
              y_i[j-1]-lam_ij_1))^(-1)*
54          (sqrt(lam_ij/lam_ij_1)*(y_i[j-1]-lam_ij_1)*((1-alpha^2)^(-3/2)))-(sqrt(lam_ij/lam_ij_1)
                *(y_i[j-1]-lam_ij_1)*((1-alpha^2)^(-3/2)))
55        l_beta_b <- l_beta_b + y_i[j]*(lam_ij+(alpha/sqrt(1-alpha^2))*sqrt(lam_ij/lam_ij_1)*(y_i[
              j-1]-lam_ij_1))^(-1)*
56          (x_i[j,]*lam_ij+(alpha/sqrt(1-alpha^2))*(0.5*sqrt(lam_ij/lam_ij_1)*(x_i[j,]-x_i[j-1,])*
                (y_i[j-1]-lam_ij_1)-x_i[j-1,]*lam_ij_1*sqrt(lam_ij/lam_ij_1)))-
57        (x_i[j,]*lam_ij+(alpha/sqrt(1-alpha^2))*(0.5*sqrt(lam_ij/lam_ij_1)*(x_i[j,]-x_i[j-1,])*(y
              _i[j-1]-lam_ij_1)-x_i[j-1,]*lam_ij_1*sqrt(lam_ij/lam_ij_1)))
58      }
59      if(j>2){
60        lam_ij <- exp(t(beta)%*%x_i[j,])
61        lam_ij <- lam_ij[1]
62        lam_ij_1 <- exp(t(beta)%*%x_i[j-1,])
63        lam_ij_1 <- lam_ij_1[1]
64        l_alpha_b <- l_alpha_b + y_i[j]*(lam_ij + alpha*sqrt(lam_ij/lam_ij_1)*(y_i[j-1]-lam_ij_1)
              )^(-1)*
65          (sqrt(lam_ij/lam_ij_1)*(y_i[j-1]-lam_ij_1))-(sqrt(lam_ij/lam_ij_1)*(y_i[j-1]-lam_ij_1))
66        l_beta_c <- l_beta_c + y_i[j]*(lam_ij + alpha*sqrt(lam_ij/lam_ij_1)*(y_i[j-1]-lam_ij_1))
              ^(-1)*
67          (x_i[j,]*lam_ij+alpha*(0.5*sqrt(lam_ij/lam_ij_1)*(x_i[j,]-x_i[j-1,])*(y_i[j-1]-lam_ij_
                1)-x_i[j-1,]*lam_ij_1*sqrt(lam_ij/lam_ij_1)))-
68        (x_i[j,]*lam_ij+alpha*(0.5*sqrt(lam_ij/lam_ij_1)*(x_i[j,]-x_i[j-1,])*(y_i[j-1]-lam_ij_1)-
              x_i[j-1,]*lam_ij_1*sqrt(lam_ij/lam_ij_1)))
69      }
70    }
71  }
72    l_beta <- l_beta_a+l_beta_b+l_beta_c
73    l_alpha <- l_alpha_a+l_alpha_b
74    score_i <- rbind(l_alpha, l_beta)
75    out.score[i,1] <- i
76    out.score[i,2:6] <- t(score_i)
77  }
78  return(out.score)
79  }
```

C.5. Supporting R functions

```
#####################################################################
## Supporting functions called in Likelihood function
## Functions written by Matthew Guerra et al. (2012):
## Guerra, M.W., Shults, J., Amsterdam, J., and Ten-Have, T. (2012).
## The analysis of binary longitudinal data with time-dependent covariates.
## Statistics in medicine 31, 931-948.
#####################################################################

cluster.size<- function(id){
  clid<- unique(id)
  m<- length(unique(id))
  n<- rep(0,m)
  autotime<- rep(0,0)
  for(i in 1:m){
    n[i]<- length(which(id==clid[i]))
    autotime<- c(autotime,1:n[i])
  }
  id<- rep(1:m,n)
  return(list(m=m,n=n,id=id,autotime=autotime))
}

data.proc<- function(data,formula,time=NULL,id,del.n){

  dat<- data.frame(data)
  col.name<- names(dat)
  #cat("1\n")
  #print(dat)

  cluster<- cluster.size(id)
  m<- cluster$m
  n<- cluster$n
  id<- cluster$id
  if(length(time)==0){
    time<- cluster$autotime
  }
  autotime<- cluster$autotime
  index<- order(id,time)
  #cat("index",index,"\n")
  #print(dat)
  #cat("ncol.dat",ncol(dat),"\n")
  if(ncol(dat)==1){
    dat<- dat[index,]
  }else{
    dat<- dat[index,]
```

```
  }
  dat<- data.frame(dat)
  names(dat)<- col.name


  del<- which(n<=del.n)
  if(length(del)>0){
    n<- n[-del]
    m<- length(n)
    mtch<- match(id,del)
    del.id<- which(mtch!="NA")
    #cat("ncol(dat)",ncol(dat),"\n")
    dat<- dat[-del.id,]
    dat<- data.frame(dat)
    names(dat)<- col.name
    row.names(dat)<- 1:nrow(dat)
    time<- time[-del.id]
    autotime<- autotime[-del.id]
    id<- rep(1:m,n)
  }

  formula<- as.formula(formula)
  fml<- as.formula(paste("~",formula[3],"+",formula[2],sep=""))
  #print(fml)
  #print(dat)
  dat<- model.matrix(fml,data=dat)

  return(list(data=dat,time=time,autotime=autotime,id=id,m=m,n=n))
}
```

C.6. R code for the epileptic seizure data application

The following appendix contains additional information to reproduce the analysis in the Application Section 4.3 for the epilepsy data. Thall and Vail, 1990 present data from a randomized, placebo-controlled study on 59 epileptic patients with seizure counts measured every 2 weeks over an 8 week period. Patients were randomized to drug treatment or placebo alongside standard chemotherapy treatment and measured the outcome as the count of the number of seizures. Additional covariates include information on patient treatment (placebo or drug), baseline seizure counts, and age in years. Of the 59 patients, 28 were randomized to placebo and 31 were randomized to drug treatment.

We begin by loading the data in the long data frame. An excerpt of the data is shown in Table C.1. The code is below.

```
#long data frame
library(MASS)
data(epil)
names(epil)
y<-epil$y
trt<-epil$trt
base<-epil$base
age<-epil$age
time<-epil$period
id<-epil$subject
```

We analyze the data using GEE and first run the model including the interaction term.

```
library(geepack)
#Run full model with interaction and get p-value of interaction
  epil_gee_int <- geeglm(y ~ trt + period + base + age + trt*period,
    data=epil, id = subject, family = poisson(link = "log"), corstr = "ar1")
  summary(epil_gee_int)
```

The relevant output of interest from the model with the interaction term is shown below.

```
Call:
geeglm(formula = y ~ trt + period + base + age + trt * period,
```

```
    family = poisson(link = "log"), data = epil, id = subject,
    corstr = "ar1")

 Coefficients:
                    Estimate  Std.err   Wald Pr(>|W|)
(Intercept)          0.55485  0.33661   2.72    0.099 .
trtprogabide        -0.10390  0.20229   0.26    0.607
period              -0.05141  0.05391   0.91    0.340
base                 0.02323  0.00124 351.21   <2e-16 ***
age                  0.02625  0.01178   4.96    0.026 *
trtprogabide:period -0.02572  0.06638   0.15    0.698
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     5.06    1.62

Correlation: Structure = ar1  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.552  0.0652
Number of clusters:   59   Maximum cluster size: 4
```

We then fit a GEE model that drops the interaction term using the following.

```
#Run reduced model without interaction
  epil_gee_main2 <- geeglm(y ~ trt + base + age + period, data=epil, id = epil$subject,
          family = poisson(link = "log"), corstr = "ar1")
  summary(epil_gee_main2)
```

The relevant output from GEE is shown below.

```
Call:
geeglm(formula = y ~ trt + base + age + period, family = poisson(link = "log"),
    data = epil, id = epil$subject, corstr = "ar1")

 Coefficients:
             Estimate  Std.err   Wald Pr(>|W|)
(Intercept)   0.58548  0.34913   2.81    0.094 .
trtprogabide -0.16422  0.15892   1.07    0.301
```

```
base           0.02322  0.00124 350.97   <2e-16 ***
age            0.02627  0.01181   4.95    0.026 *
period        -0.06445  0.03400   3.59    0.058 .
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     5.07    1.64

Correlation: Structure = ar1  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.551  0.0656
Number of clusters:   59   Maximum cluster size: 4
```

Lastly, we run the GEE model with treatment, baseline seizure count, and age only removing the time variable (period) from the model.

```
#Remove time variable from model
  epil_gee_main1 <- geeglm(y ~ trt + base + age, data=epil, id = epil$subject,
          family = poisson(link = "log"), corstr = "ar1")
  summary(epil_gee_main1)
```

The relevant output is shown below.

```
Call:
geeglm(formula = y ~ trt + base + age, family = poisson(link = "log"),
    data = epil, id = epil$subject, corstr = "ar1")

 Coefficients:
             Estimate  Std.err   Wald Pr(>|W|)
(Intercept)   0.44670  0.36212   1.52    0.217
trtprogabide -0.16588  0.15928   1.08    0.298
base           0.02316  0.00123 353.32   <2e-16 ***
age            0.02576  0.01169   4.86    0.028 *
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1

Estimated Scale Parameters:
```

```
             Estimate Std.err
(Intercept)     5.07    1.57

Correlation: Structure = ar1  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.544  0.0639
Number of clusters:   59   Maximum cluster size: 4
```

Next, we analyze the epilepsy data using the ML approach described. For demonstrating the code in detail, we begin with the model with the main effects of treatment, baseline seizure counts, and age. After loading the data, we ran the GEE model with these main effects and will use these starting values. The output from the GEE model is shown above and includes the vector of starting values to be used for the ML approach.

```
# Step 1: Run the epil data (see above)
# Step 2: Obtain starting values by fitting a GEE model
# Step 3: Set starting values
beta.start <- epil_gee_main1$geese$beta
alpha.start <- epil_gee_main1$geese$alpha
start.values <- t(t(c(alpha.start,beta.start)))
```

After checking that the constraints defined in Section 4.2.2 are met, we next define the starting values. Consider the vector of parameters to be defined as $\theta = (\alpha, \beta_0, \beta_1, \beta_2, \beta_3)^T$. Our algorithm uses starting values from the function `geeglm` in the package `geepack`. The R code for the model and starting value assignment would be given by

```
model <- geeglm(y ~ x1 + x2 + x3, data=indata, id = indata$subject, family = poisson(link = "
    log"), corstr = "ar1")
beta.start <- model$geese$beta
alpha.start <- model$geese$alpha
start.values <- t(t(c(alpha.start,beta.start)))
```

Next, we define the feasibility region constraints to run the `constrOptim` function for the vector of parameters for the linear predictor. For the epilepsy data, we implement Model 4.5 with the following

linear predictor:

$$x'_{ij}\beta = \beta_0 + \beta_1 x_{ij1} + \beta_2 x_{ij2} + \beta_3 x_{ij3}, \tag{C.7}$$

where x_{ij1} represents an indicator for treatment, x_{ij2} represents baseline seizure count, and x_{ij3} represents subject age.

Define `ui` as the $k \times p$ constraint matrix and `ci` as the constraint vector of length k. Then, the linear inequality constraints, or feasibility region, is defined by `ui`$\% * \%$ `theta` $-$ `ci`≥ 0 where the $p \times 1$ vector of parameters is represented by `theta`. This vector of parameters is also an argument in the function representing the numeric $p \times 1$ vector of initial values, which must be on the interior of the earlier defined feasibility region.

In matrix notation, for a model with parameters $\theta = (\alpha, \beta_0, \beta_1, \beta_2, \beta_3)^T$, this constraint is given by

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{pmatrix} - \begin{pmatrix} -1 \\ -1 \end{pmatrix} \geq \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Then, we get $\alpha + 1 \geq 0$ and $-\alpha + 1 \geq 0$, , which can be written as $\alpha \geq -1$ and as $\alpha \leq 1$, respectively. In R, the feasibility region is specified by

```
#constraints
ui <- rbind(c(1,0,0,0,0), c(-1,0,0,0,0))
ci <- c(-1,-1)
full.ml <- constrOptim(start.values, logl3, grad=ml.grad, ui = ui, ci = ci, mu = 1e-04, control
    =list("fnscale"=-1), outer.iterations = 100, outer.eps = 1e-05, hessian = TRUE)
```

With the log-likelihood function written and input into the constrained maximization, we check the constraint $\lambda_{ij} - \alpha\sigma_{ij}/\sigma_{ij-1}(\lambda_{ij-1}) > 0$, for $(j = 2, \ldots, n_i)$. This is evaluated within the log likelihood function when determining values for λ^*_{ij}, where $j = 2, \ldots, n_i$.

Refer to Appendix C.2 for the log likelihood function for this data and Appendix C.3 for the gradient function. These functions each take the vector of starting values and should be programmed out-

side of the function call. This allows the specification `hessian=TRUE` to be used, which is necessary in order to return the numerical Hessian matrix and is used to obtain the estimated covariance matrix of the maximum likelihood estimates. As noted previously, since the log likelihood is provided, the option $\texttt{control} = \texttt{list}(\text{``fnscale''} = -1)$ must be added to maximize the objective function. Both the objective function `f` and the gradient function `grad` take the argument `theta`, the vector of parameters over which the maximization is to take place. The function `f` returns a scalar result while the function `grad` returns the gradient for the BFGS method.

We specify the maximum number of iterations of the barrier algorithm to 100 (in the code this is given by `outer.iterations = 100`) and the relative convergence tolerance of the barrier algorithm to 0.00001 (in the code this is given by `outer.eps = 1e-05`) for the `constrOptim` function to run.

The output from the starting values and the constrained optimization for the maximum likelihood approach are shown below.

```
start.values
                  [,1]
alpha          0.5443
(Intercept)    0.4467
trtprogabide  -0.1659
base           0.0232
age            0.0258

> full.ml
$par
                  [,1]
alpha          0.4227
(Intercept)    0.5072
trtprogabide  -0.1673
base           0.0232
age            0.0238

$value
[1] -781

$counts
function gradient
      60       10

$convergence
[1] 0

```

```
$message
NULL

$hessian
         [,1]    [,2]      [,3]      [,4]      [,5]
[1,]   -902.9     211      97.6     13626      6272
[2,]    210.6    -999    -501.4    -62587    -27916
[3,]     97.6    -501    -501.4    -35825    -13040
[4,] 13626.4 -62587 -35825.0 -6372318 -1628104
[5,]  6272.2 -27916 -13039.8 -1628104  -820673

$outer.iterations
[1] 2

$barrier.value
[1] 1.84e-05
```

We compute the AIC and BIC for this model.

```
mle.beta <- full.ml$par[2:5]
mle.alpha <- full.ml$par[1]
# log likelihood:
mle.full <- full.ml$value
AIC <- 2*(length(mle.beta)+1)-2*(mle.full)
BIC <- log(length(unique(id)))*(length(mle.beta)+1)-2*(mle.full)
> AIC
[1] 1573
> BIC
[1] 1583
```

For the score estimation and the covariance matrix, we use the following code.

```
# covariance matrix from the constrOptim output
mle.cov <- solve(-full.ml$hessian)
#returns matrix for score of each subject i=1,m in rows
out.score <- try(score_squared(formula=formula, data=data, id=id, time=time, alpha=mle.alpha,
    beta=t(mle.beta)))
p <- length(mle.beta) + 1
score.sq <- matrix(0, nrow=p, ncol=p)
for (i in 1:m){
  score.sq <- score.sq + out.score[i,2:6]%*%t(out.score[i,2:6])
}
```

```
sq_cov <- (solve(score.sq))

#ob = observed information = 1/i(hat(theta))
std.err <- "ob"
if (std.err=="ob"){
  mle_cov <- mle.cov
}
```

```
# Observed information
> mle_cov
           [,1]       [,2]       [,3]       [,4]       [,5]
[1,]  1.17e-03 -1.95e-04  1.08e-07  8.88e-07  1.38e-05
[2,] -1.95e-04  3.59e-02 -4.78e-03 -6.72e-05 -1.01e-03
[3,]  1.08e-07 -4.78e-03  4.44e-03 -3.13e-06  9.81e-05
[4,]  8.88e-07 -6.72e-05 -3.13e-06  4.83e-07  1.38e-06
[5,]  1.38e-05 -1.01e-03  9.81e-05  1.38e-06  3.15e-05
```

Lastly, we conduct the hypothesis testing using either the estimated covariance matrix.

```
Stderr <- matrix(NA, nrow=pp, ncol=1)
Wald <- matrix(NA, nrow=pp, ncol=1)
pval <- matrix(NA, nrow=pp, ncol=1)
for (p in 1:pp){
  Stderr[p,] <- sqrt(mle_cov[(p+1),(p+1)])
  Wald[p,] <- (mle.beta[p]/sqrt(mle_cov[(p+1),(p+1)]))^2
  pval[p,] <- 1-pchisq(Wald[p,1] , df=1, lower.tail = TRUE, log.p = FALSE)
}
results <- cbind(mle.beta,Stderr, Wald, pval)
alpha_results <- cbind(mle.alpha,sqrt(mle_cov[1,1]))
```

The output from the hypothesis testing is shown below.

```
> print(results)
     Estimate  Std.err    Wald Pr(>|W|)
[1,]   0.5072 0.189399    7.17 7.41e-03
[2,]  -0.1673 0.066661    6.30 1.21e-02
[3,]   0.0232 0.000695 1113.57 0.00e+00
[4,]   0.0238 0.005609   17.99 2.22e-05
> alpha_results
      Estimate Std.err
```

```
alpha    0.423  0.0342
```

Next, we fit the model with period included.

```
#Run reduced model without interaction and get loglik
#y ~ trt + period + base + age

beta.start3 <- epil_gee_main2$geese$beta
alpha.start3 <- epil_gee_main2$geese$alpha
start.values3 <- t(t(c(alpha.start3,beta.start3)))

#constraints
ui <- rbind(c(1,0,0,0,0,0), c(-1,0,0,0,0,0))
ci <- c(-1,-1)

ml.full.time <- constrOptim(start.values3, logL.time, grad=ml.grad.time, ui = ui, ci = ci, mu =
     1e-04, control=list("fnscale"=-1), outer.iterations = 100, outer.eps = 1e-05, hessian =
    TRUE)
```

The relevant output from this model is shown below.

```
> ml.full.time
$par
                 [,1]
alpha          0.4159
(Intercept)    0.6569
trtprogabide -0.1661
base         -0.0635
age            0.0232
period         0.0238

$value
[1] -777

$counts
function gradient
     120       17

$convergence
[1] 0

$message
```

```
NULL

$hessian
         [,1]    [,2]      [,3]     [,4]      [,5]      [,6]
[1,]  -934.6     163      74.8      476     10479      4939
[2,]   162.9    -998    -501.8    -2286    -62510    -27905
[3,]    74.8    -502    -501.8    -1150    -35796    -13052
[4,]   476.1   -2286   -1150.0    -7412   -142770    -63939
[5,] 10479.0 -62510 -35795.7 -142770 -6361288 -1625952
[6,]  4939.2 -27905 -13052.5   -63939 -1625952   -820229

$outer.iterations
[1] 2

$barrier.value
[1] 1.78e-05
```

Using this output, we compute the AIC, BIC, and model statistics as described in detail for the previous model.

```
##Fit statistics
 formula <- y ~ trt + period + base + age

  mle.beta <- ml.full.time$par[2:6]
  mle.alpha <- ml.full.time$par[1]
  mle.full <- ml.full.time$value #this is the log likelihood
  mle.cov <- solve(-ml.full.time$hessian) #covariance matrix

  AIC <- 2*(length(mle.beta)+1)-2*(mle.full)
  BIC <- log(length(unique(id)))*(length(mle.beta)+1)-2*(mle.full)
  pp <- length(all.vars(formula))


#Score estimation:
  out.score <- try(score_squared.time(formula=formula, data=data, id=id, time=time,
            mle.alpha, t(mle.beta)))
    #returns matrix for score of each subject i=1,m in rows
  p <- length(mle.beta) + 1
  score.sq <- matrix(0, nrow=p, ncol=p)
  for (i in 1:m){
    score.sq <- score.sq + out.score[i,2:7]%*%t(out.score[i,2:7])
  }
  sq_cov <- (solve(score.sq))

```

```
  std.err<-"ob"
  if (std.err=="ob"){
    mle_cov <- mle.cov
    }

## Hypothesis testing
  Stderr <- matrix(NA, nrow=pp, ncol=1)
  Wald <- matrix(NA, nrow=pp, ncol=1)
  pval <- matrix(NA, nrow=pp, ncol=1)
  for (p in 1:pp){
    Stderr[p,] <- sqrt(mle_cov[(p+1),(p+1)])
    Wald[p,] <- (mle.beta[p]/sqrt(mle_cov[(p+1),(p+1)]))^2
    pval[p,] <- 1-pchisq(Wald[p,1] , df=1, lower.tail = TRUE, log.p = FALSE)
  }
  results <- cbind(mle.beta,Stderr, Wald, pval)
  alpha_results <- cbind(mle.alpha,sqrt(mle_cov[1,1]))
  fit_stats <- rbind(mle.full,AIC,BIC)

  #format output
  rownames(fit_stats) <- c("Log-Likelihood:", "AIC:", "BIC:")
  colnames(results) <- c("Estimate", "Std.err", "Wald", "Pr(>|W|)")
  rownames(results) <- c("(Intercept)", "trt", "period", "base", "age")
  colnames(alpha_results) <- c("Estimate", "Std.err")
  rownames(alpha_results) <- c("alpha") #only if corr=="ar1"

  print(fit_stats)
  print(results)
  print(alpha_results)
```

The output is shown below.

```
> print(fit_stats)
                      [,1]
Log-Likelihood: -777.10421
AIC:            1566.20842
BIC:            1578.67365
> print(results)
             Estimate  Std.err          Wald
(Intercept)  0.65688  0.19575    11.261174
trt         -0.16611  0.06666     6.210783
period      -0.06352  0.02149     8.736366
base         0.02317  0.00069  1111.759676
age          0.02376  0.00561    17.937422
               Pr(>|W|)
```

```
(Intercept) 7.9145e-04
trt         1.2697e-02
period      3.1193e-03
base        0.0000e+00
age         2.2829e-05
> print(alpha_results)
      Estimate  Std.err
alpha 0.41587  0.03336
```

Here, the AIC is approximately 1566.208 and the BIC is approximately 1578.674.

We then do a likelihood ratio test for the time variable.

```
#LR Test Statistic
#G = -2*(logL(reduced) - logL(full))
G.time <- -2*( ml.full.notime$value - ml.full.time$value)
pval.time <- 1-pchisq(G.time , df=1, lower.tail = TRUE, log.p = FALSE)
#assuming reduced model (null) is correct, the sampling distr of G is approx
#  chi-squared with df=1 if only 1 interaction term
```

Next, we fit the model with the interaction term.

```
#Run full model with interaction and get loglik

beta.start2 <- epil_gee_int$geese$beta
alpha.start2 <- epil_gee_int$geese$alpha
start.values <- t(t(c(alpha.start2,beta.start2)))

#constraints
ui <- rbind(c(1,0,0,0,0,0,0), c(-1,0,0,0,0,0,0))
ci <- c(-1,-1)

ml.full.int <- constrOptim(start.values, logL.int, grad=ml.grad.int, ui = ui, ci = ci, mu = 1e
    -04, control=list("fnscale"=-1), outer.iterations = 100, outer.eps = 1e-05, hessian = TRUE
    )
```

The relevant output is shown below.

```
> ml.full.int
```

```
$par
                         [,1]
alpha                  0.4162
(Intercept)            0.6569
trtprogabide          -0.1225
period                -0.0634
base                   0.0232
age                    0.0238
trtprogabide:period   -0.0443

$value
[1] -777

$counts
function gradient
       61       10

$convergence
[1] 0

$message
NULL

$hessian
          [,1]    [,2]      [,3]     [,4]      [,5]      [,6]      [,7]
[1,]   -935.2     163      74.6      477     10493      4948      74.6
[2,]    163.2    -998    -501.3    -2285    -62476    -27891    -501.3
[3,]     74.6    -501    -501.3    -1149    -35763    -13040    -501.3
[4,]    477.0   -2285   -1148.8    -7409   -142687    -63905   -1148.8
[5,] 10492.9 -62476 -35762.9 -142687 -6357559 -1625171 -35762.9
[6,]  4947.7 -27891 -13039.7  -63905 -1625171  -819850 -13039.7
[7,]    74.6    -501    -501.3    -1149    -35763    -13040    -501.3

$outer.iterations
[1] 2

$barrier.value
[1] 1.79e-05
```

As shown above, the following formulas are used to compute the AIC and BIC.

```
AIC <- 2*(length(mle.beta)+1)-2*(mle.full)
BIC <- log(length(unique(id)))*(length(mle.beta)+1)-2*(mle.full)
```

For the model with the interaction term, the AIC is 1568.208 and the BIC is 1582.751.

The code to conduct the likelihood ratio test is shown below.

```
> options(digits=20)
> ml.full.time$value
[1] -777.10421068667892541
> ml.full.int$value
[1] -777.10409476684776564
#LR Test Statistic
#G = -2*(logL(reduced) - logL(full))
G.int <- -2*( ml.full.time$value - ml.full.int$value)
pval.int <- 1-pchisq(G.int , df=1, lower.tail = TRUE, log.p = FALSE)

> G.int
[1] 0.00023183966231954400428
> pval.int
[1] 0.98785165411905362
```

C.7. R code for the doctor visits data application

The following appendix contains additional information to reproduce the analyses in the Application Section 4.3.1 for the doctor visits data. We analyzed a subset of data from the German Socio-Economic Panel data (Winkelmann, 2004) (`http://www.stata-press.com/data/r13/drvisits`) that we obtained within Stata and then exported as a comma delimited text-file for analysis in R (StataCorp LP, 2013). Here we compare the results of an analysis using the proposed ML approach with the results obtained using Poisson regression and GEE.

The goal of the analysis was to assess the impact of the 1997 health reform on the reduction of government expenditures. A sample of 1518 women who were employed full time in the year before or in the year after the reform was used to assess the impact on the number of doctor visits. The outcome was the self-reported number of doctor visits in the most recent three months prior to the interview. The main covariate of interest was the indicator of whether the interview was before the reform or after the reform and covariate information was available on the women's age, education, marital status, self-reported health status, and the logarithm of the household income.

We begin by loading the data into R.

```
1 drvisits_raw <- read.table("C:/Users/Victoria/Desktop/Post_Mtg_20150623/drvisits.raw", header=
      TRUE, sep=",")
2 summary(drvisits_raw)
```

Next, we run the Poisson regression using the following code.

```
1 drv_poi <- glm(numvisit ~ reform + age + educ + married + badh + loginc, data=drvisits_raw,
      family = poisson)
2 summary(drv_poi)
```

The relevant output is shown below.

```
1 > summary(drv_poi)
2
3 Call:
4 glm(formula = numvisit ~ reform + age + educ + married + badh +
5     loginc, family = poisson, data = drvisits_raw)
```

```

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-3.963   -1.934   -0.672    0.550   12.659

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.41367    0.26909   -1.54   0.1242
reform      -0.14015    0.02655   -5.28  1.3e-07 ***
age          0.00437    0.00130    3.35   0.0008 ***
educ        -0.01072    0.00601   -1.78   0.0743 .
married      0.04135    0.02784    1.49   0.1375
badh         1.13317    0.03030   37.40  < 2e-16 ***
loginc       0.14923    0.03605    4.14  3.5e-05 ***
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 8848.8  on 2226  degrees of freedom
Residual deviance: 7419.9  on 2220  degrees of freedom
AIC: 11899

Number of Fisher Scoring iterations: 5
```

Next, we use GEE to analyze this data.

```
drv_gee <- geeglm(numvisit ~ reform + age + educ + married + badh + loginc, data=drvisits_raw,
     id = id,family = poisson(link = "log"), corstr = "ar1")
summary(drv_gee)
```

The relevant output is shown below.

```
> summary(drv_gee)

Call:
geeglm(formula = numvisit ~ reform + age + educ + married + badh +
    loginc, family = poisson(link = "log"), data = drvisits_raw,
    id = id, corstr = "ar1")

 Coefficients:
             Estimate  Std.err   Wald Pr(>|W|)
```

```
(Intercept) -0.38146  0.57665   0.44    0.508
reform      -0.12300  0.05295   5.40    0.020 *
age          0.00522  0.00334   2.44    0.118
educ        -0.00920  0.01178   0.61    0.435
married      0.03842  0.06983   0.30    0.582
badh         1.10549  0.08733 160.23   <2e-16 ***
loginc       0.13920  0.07976   3.05    0.081 .
---
Signif. codes:  0    ***    0.001    **    0.01    *    0.05    .    0.1         1

Estimated Scale Parameters:
            Estimate Std.err
(Intercept)     4.33   0.369

Correlation: Structure = ar1  Link = identity

Estimated Correlation Parameters:
      Estimate Std.err
alpha    0.213  0.0238
Number of clusters:   1518   Maximum cluster size: 2
```

After checking that the constraints defined in Section 4.2.2 are met, we next define the starting values from GEE that are used for the ML approach.

```
alpha <- start.values[1]
 beta <- start.values[2:length(start.values)]

#to be updated by user:
 formula <- numvisit ~ reform + age + educ + married + badh + loginc
 id <- drvisits_raw$id
 time <- drvisits_raw$visit
 d <- dim(drvisits_raw)
 k <- length(all.vars(formula))-1
 dt.fm<- data.frame(drvisits_raw)
```

We then specify the feasibility region and call the ML approach.

```
#constraints
ui <- rbind(c(1,0,0,0,0,0,0,0), c(-1,0,0,0,0,0,0,0))
ci <- c(-1,-1)

```

```
full.ml2 <- constrOptim(start.values, logl3, grad=ml.grad, ui = ui, ci = ci, mu = 1e-04,
    control=list("fnscale"=-1), outer.iterations = 100, outer.eps = 1e-05, hessian = TRUE)
full.ml2
```

The relevant output is shown below.

```
$par
                  [,1]
alpha         0.313022
(Intercept) -0.461286
reform       -0.113059
age           0.004914
educ         -0.007953
married       0.025525
badh          1.100140
loginc        0.149758

$value
[1] -5845

$counts
function gradient
      69       12

$convergence
[1] 0

$message
NULL

$hessian
         [,1]       [,2]     [,3]     [,4]     [,5]       [,6]     [,7]     [,8]
[1,] -2490.4     422.9    539.2    17021     5150     145.9   -145.4     3266
[2,]   422.9   -4911.4  -2182.6  -188698   -55905   -2520.7  -1484.8   -37923
[3,]   539.2   -2182.6  -2754.6   -85547   -25212   -1148.6   -570.8   -16878
[4,] 17020.7 -188698.1 -85547.3 -7862830 -2126640 -105187.7 -64147.6 -1458400
[5,]  5150.3  -55904.9 -25211.7 -2126640  -663877  -27784.6 -16238.9  -432853
[6,]   145.9   -2520.7  -1148.6  -105188   -27785   -2583.1   -857.1   -19501
[7,]  -145.4   -1484.8   -570.8   -64148   -16239    -857.1  -1597.5   -11417
[8,]  3265.9  -37923.3 -16877.7 -1458400  -432853  -19500.7 -11417.1  -293595

$outer.iterations
[1] 2

```

```
$barrier.value
[1] 9.965e-06
```

Using this information we compute the score function and the observed information.

```
out.score <- try(score_squared(formula=formula, data=data, id=id, time=time, mle.alpha=mle.
    alpha, mle.beta=mle.beta))
  #returns matrix for score of each subject i=1,m in rows
p <- length(mle.beta) + 1
score.sq <- matrix(0, nrow=p, ncol=p)
for (i in 1:m){
  score.sq <- score.sq + out.score[i,2:9]%*%t(out.score[i,2:9])
}
sq_cov <- (solve(score.sq))
std.err <- "ob"
if (std.err=="ob"){
  mle_cov <- mle.cov
}
```

The covariance matrix is shown below.

```
> mle_cov
            [,1]        [,2]        [,3]        [,4]        [,5]        [,6]
[1,]  4.324e-04  5.686e-05  7.788e-05  2.093e-06  3.039e-06 -2.870e-05
[2,]  5.686e-05  7.901e-02 -8.847e-05 -3.724e-05 -1.886e-05  1.532e-04
[3,]  7.788e-05 -8.847e-05  5.812e-04 -2.024e-06 -7.919e-06 -1.219e-05
[4,]  2.093e-06 -3.724e-05 -2.024e-06  1.976e-06  1.171e-06 -1.102e-05
[5,]  3.039e-06 -1.886e-05 -7.919e-06  1.171e-06  4.119e-05  2.254e-05
[6,] -2.870e-05  1.532e-04 -1.219e-05 -1.102e-05  2.254e-05  8.652e-04
[7,] -1.063e-04 -4.929e-04  3.548e-05 -1.151e-05  1.058e-05  1.310e-05
[8,] -1.585e-05 -9.978e-03  4.273e-08 -5.411e-06 -6.552e-05 -5.584e-05
            [,7]        [,8]
[1,] -1.063e-04 -1.585e-05
[2,] -4.929e-04 -9.978e-03
[3,]  3.548e-05  4.273e-08
[4,] -1.151e-05 -5.411e-06
[5,]  1.058e-05 -6.552e-05
[6,]  1.310e-05 -5.584e-05
[7,]  9.774e-04  6.313e-05
[8,]  6.313e-05  1.417e-03
```

Next, we conduct the hypothesis testing using the observed information matrix.

```
Stderr <- matrix(NA, nrow=pp, ncol=1)
Wald <- matrix(NA, nrow=pp, ncol=1)
pval <- matrix(NA, nrow=pp, ncol=1)
for (p in 1:pp){
  Stderr[p,] <- sqrt(mle_cov[(p+1),(p+1)])
  Wald[p,] <- (mle.beta[p]/sqrt(mle_cov[(p+1),(p+1)]))^2
  pval[p,] <- 1-pchisq(Wald[p,1] , df=1, lower.tail = TRUE, log.p = FALSE)
}
results <- cbind(mle.beta,Stderr, Wald, pval)
alpha_results <- cbind(mle.alpha,sqrt(mle_cov[1,1]))
colnames(results) <- c("Estimate", "Std.err", "Wald", "Pr(>|W|)")
rownames(results) <- c("(Intercept)", "reform", "age", "educ", "married", "badh", "loginc")
colnames(alpha_results) <- c("Estimate", "Std.err")
rownames(alpha_results) <- c("alpha") #only if corr=="ar1"
fit_stats <- rbind(mle.full,AIC,BIC)
print(fit_stats)
print(results)
print(alpha_results)
```

The relevant output is shown below.

```
> print(fit_stats)
                  [,1]
Log-Likelihood: -5845
AIC:            11707
BIC:            11750
> print(results)
             Estimate  Std.err     Wald  Pr(>|W|)
(Intercept) -0.461286 0.281089    2.693 1.008e-01
reform      -0.113059 0.024107   21.994 2.735e-06
age          0.004914 0.001406   12.220 4.727e-04
educ        -0.007953 0.006418    1.536 2.153e-01
married      0.025525 0.029415    0.753 3.855e-01
badh         1.100140 0.031264 1238.276 0.000e+00
loginc       0.149758 0.037641   15.829 6.932e-05
> print(alpha_results)
      Estimate Std.err
alpha    0.313 0.02079
```

Lastly, we include the code to conduct a likelihood ratio test for the correlation parameter. Note that

the log-likelihood values are obtained from the ML and Poisson models above.

```
1
2 #number estimated parameters in Poisson (p.noalpha) vs ML (p.alpha)
3 p.alpha <- 8
4 p.noalpha <- 7
5
6 #log likelihood in Poisson (p.noalpha) vs ML (p.alpha)
7 ml.noalpha <- -5942.5
8 ml.alpha <- -5845.5
9
10 #LR Test Statistic
11 #G = -2*(logL(reduced) - logL(full))
12 G.alpha <- -2*( ml.noalpha - ml.alpha)
13 pval.alpha <- 1-pchisq(G.alpha , df=1, lower.tail = TRUE, log.p = FALSE)
14 #assuming reduced model (null) is correct, the sampling distr of G is approx
15 #  chi-squared with df=1 if only 1 interaction term
16 > G.alpha
17 [1] 194
18 > pval.alpha
19 [1] 0
```

Table C.1: An excerpt of the data from a randomized, placebo-controlled study on 59 epileptic patients with seizure counts measured every 2 weeks over an 8 week period (Thall and Vail, 1990).

	ID	Week 2	Week 4	Week 6	Week 8	Baseline	Age	Treatment
1	104	5	3	3	3	11	31	placebo
2	106	3	5	3	3	11	30	placebo
3	107	2	4	0	5	6	25	placebo
4	114	4	4	1	4	8	36	placebo
5	116	7	18	9	21	66	22	placebo
6	118	5	2	8	7	27	29	placebo
7	123	6	4	0	2	12	31	placebo
8	126	40	20	23	12	52	42	placebo
9	130	5	6	6	5	23	37	placebo
10	135	14	13	6	0	10	28	placebo
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
29	101	11	14	9	8	76	18	progabide
30	102	8	7	9	4	38	32	progabide
31	103	0	4	3	0	19	20	progabide
32	108	3	6	1	3	10	30	progabide
33	110	2	6	7	4	19	18	progabide
34	111	4	3	1	3	24	24	progabide
35	112	22	17	19	16	31	30	progabide
36	113	5	4	7	4	14	35	progabide
37	117	2	4	0	4	11	27	progabide
38	121	3	7	7	7	67	20	progabide
39	122	4	18	2	5	41	22	progabide
40	124	2	1	1	0	7	28	progabide
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮

BIBLIOGRAPHY

Akaike, H (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19, 716–723.

American Cancer Society (2006). *American Cancer Society: Cancer Facts and Figures 2006*. American Cancer Society.

Balch, CM (1992). "Cutaneous melanoma: prognosis and treatment results worldwide". In: *Seminars in Surgical Oncology*. Vol. 8. Wiley Online Library, 400–414.

Balch, CM, Buzaid, AC, Soong, SJ, Atkins, MB, Cascinelli, N, Coit, DG, Fleming, ID, Gershenwald, JE, Houghton, A, Kirkwood, JM, et al. (2001). Final version of the American Joint Committee on Cancer staging system for cutaneous melanoma. *Journal of Clinical Oncology* 19, 3635–3648.

Balch, CM, Gershenwald, JE, Soong, SJ, Thompson, JF, Atkins, MB, Byrd, DR, Buzaid, AC, Cochran, AJ, Coit, DG, Ding, S, et al. (2009). Final version of 2009 AJCC melanoma staging and classification. *Journal of Clinical Oncology* 27, 6199–6206.

Balch, CM, Mihm, M, Gersenwald, J, and Soong, SJ (2010). The revised melanoma staging system and the impact of mitotic rate. *The Melanoma Letter* 28.

Barchielli, A, Paci, E, Balzi, D, Geddes, M, Giorgi, D, Zappa, M, Bianchi, S, and Buiatti, E (1994). Population-based breast cancer survival. *Cancer* 74, 3126–3134.

Bennette, C and Vickers, A (2012). Against quantiles: categorization of continuous variables in epidemiologic research, and its discontents. *BMC Medical Research Methodology* 12, 21.

Broyden, CG (1970). The convergence of a class of double-rank minimization algorithms 1. general considerations. *Journal of the Institute of Mathematics and Its Applications* 6, 76–90.

Bryant, H, Lockwood, G, Rahal, R, and Ellison, L (2012). Conditional survival in Canada: adjusting patient prognosis over time. *Current Oncology* 19, 222–224.

Choi, M, Fuller, CD, Thomas Jr, CR, and Wang, SJ (2008). Conditional survival in ovarian cancer: results from the SEER dataset 1988–2001. *Gynecologic Oncology* 109, 203–209.

Consul, P and Famoye, F (1992). Generalized Poisson regression model. *Communications in Statistics-Theory and Methods* 21, 89–109.

Corbière, F and Joly, P (2007). A SAS macro for parametric and semiparametric mixture cure models. *Computer Methods and Programs in Biomedicine* 85, 173–180.

Dennis, LK (1999). Analysis of the melanoma epidemic, both apparent and real: data from the 1973 through 1994 surveillance, epidemiology, and end results program registry. *Archives of Dermatology* 135, 275–280.

deVries, E, Nijsten, TE, Visser, O, Bastiaannet, E, vanHattem, S, Janssen-Heijnen, ML, and Coebergh, JW (2008). Superior survival of females among 10 538 Dutch melanoma patients is independent of Breslow thickness, histologic type and tumor site. *Annals of Oncology* 19, 583–589.

Dickman, PW and Adami, HO (2006). Interpreting trends in cancer patient survival. *Journal of Internal Medicine* 260, 103–117.

Dickman, PW, Sloggett, A, Hills, M, and Hakulinen, T (2004). Regression models for relative survival. *Statistics in Medicine* 23, 51–64.

Ederer, F, Axtell, LM, and Cutler, SJ (1961). The relative survival rate: A statistical methodology. *National Cancer Institute Monograph* 6, 101–121.

Efron, B (1992). Poisson overdispersion estimates based on the method of asymmetric maximum likelihood. *Journal of the American Statistical Association* 87, 98–107.

Efron, B and Hinkley, DV (1978). Assessing the accuracy of the maximum likelihood estimator: Observed versus expected Fisher information. *Biometrika* 65, 457–483.

Ellison, LF, Bryant, H, Lockwood, G, and Shack, L (2011). Conditional survival analyses across cancer sites. *Health Reports* 22, 21–25.

Famoye, F, Okafor, R, and Adamu, M (2011). A multivariate generalized Poisson distribution. *Journal of Statistical Theory and Applications* 10, 519–531.

Farewell, DM and Farewell, VT (2012). Dirichlet negative multinomial regression for overdispersed correlated count data. *Biostatistics* 14, 395–404.

Fitzmaurice, G, Davidian, M, Verbeke, G, and Molenberghs, G (2008). *Longitudinal Data Analysis*. CRC Press.

Fletcher, R (1970). A new approach to variable metric algorithms. *The Computer Journal* 13, 317–322.

Gabriel, KR (1962). Ante-dependence analysis of an ordered set of variables. *The Annals of Mathematical Statistics*, 201–212.

Gamel, JW and Vogel, RL (1997). Comparison of parametric and non-parametric survival methods using simulated clinical data. *Statistics in Medicine* 16, 1629–1643.

Gamerman, V, Karakousis, GC, Guerry, D, and Gimotty, PA (2012). "Conditional survival (CS) in patients with stage II melanoma." In: *ASCO Annual Meeting Proceedings*. Vol. 30. 15_suppl.

Gardiner, JC, Luo, Z, and Roman, LA (2009). Fixed effects, random effects and GEE: what are the differences? *Statistics in Medicine* 28, 221–239.

Gimotty, PA, Botbyl, J, Soong, SJ, and Guerry, D (2005). A population-based validation of the American Joint Committee on Cancer melanoma staging system. *Journal of Clinical Oncology* 23, 8065–8075.

Goldfarb, D (1970). A family of variable-metric methods derived by variational means. *Mathematics of Computation* 24, 23–26.

Greenland, S (1995). Avoiding power loss associated with categorization and ordinal scores in dose-response and trend analysis. *Epidemiology* 6, 450–454.

Greenwood, M (1926). A report on the natural duration of cancer. *Reports on Public Health and Medical Subjects* 33, 23–25.

Grizzle, JE, Starmer, CF, and Koch, GG (1969). Analysis of categorical data by linear models. *Biometrics*, 489–504.

Guerra, MW and Shults, J (2014). A note on the simulation of overdispersed random variables with specified marginal means and product correlations. *The American Statistician* 68, 104–107.

Guerra, MW, Shults, J, Amsterdam, J, and Ten-Have, T (2012). The analysis of binary longitudinal data with time-dependent covariates. *Statistics in Medicine* 31, 931–948.

Hakulinen, T (1982). Cancer survival corrected for heterogeneity in patient withdrawal. *Biometrics* 38, 933–942.

Heagerty, PJ and Kurland, BF (2001). Misspecified maximum likelihood estimates and generalised linear mixed models. *Biometrika* 88, 973–985.

Hieke, S, Kleber, M, König, C, Engelhardt, M, and Schumacher, M (2015). Conditional survival: A useful concept to provide information on how prognosis evolves over time. *Clinical Cancer Research* 21, 1530–1536.

Hutton, J and Monaghan, P (2002). Choice of parametric accelerated life and proportional hazards models for survival data: Asymptotic results. *Lifetime Data Analysis* 8, 375–393.

Jeong, JH, Jung, SH, and Costantino, JP (2008). Nonparametric inference on median residual life function. *Biometrics* 64, 157–163.

Jung, SH, Jeong, JH, and Bandos, H (2009). Regression on quantile residual life. *Biometrics* 65, 1203–1212.

Kaplan, EL and Meier, P (1958). Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association* 53, 457–481.

Klein, JP and Moeschberger, ML (2005). *Survival analysis: techniques for censored and truncated data*. Springer Science & Business Media.

Koch, GG, Johnson, WD, and Tolley, HD (1972). A linear models approach to the analysis of survival and extent of disease in multidimensional contingency tables. *Journal of the American Statistical Association* 67, 783–796.

Lachin, JM (2000). *Biostatistical methods: The assessment of relative risks*. John Wiley & Sons.

Lambert, PC and Royston, P (2009). Further development of flexible parametric models for survival analysis. *Stata Journal* 2, 265–290.

Liang, KY and Zeger, SL (1986). Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13–22.

MacKie, RM, Hole, D, Hunter, JA, Rankin, R, Evans, A, McLaren, K, Fallowfield, M, Hutcheon, A, and Morris, A (1997). Cutaneous malignant melanoma in Scotland: incidence, survival, and mortality, 1979-94. *British Medical Journal* 315, 1117–1121.

Merrill, RM, Henson, DE, and Ries, LA (1998). Conditional survival estimates in 34,963 patients with invasive carcinoma of the colon. *Diseases of the Colon & Rectum* 41, 1097–1106.

Merrill, RM and Hunter, BD (2010). Conditional survival among cancer patients in the United States. *The Oncologist* 15, 873–882.

Miller, BJ, Lynch, CF, and Buckwalter, JA (2013). Conditional survival is greater than overall survival at diagnosis in patients with osteosarcoma and Ewings sarcoma. *Clinical Orthopaedics and Related Research* 471, 3398–3404.

Molenberghs, G and Kenward, MG (2010). Semi-parametric marginal models for hierarchical data and their corresponding full models. *Computational Statistics & Data Analysis* 54, 585–597.

Morton, DL, Thompson, JF, Cochran, AJ, Mozzillo, N, Elashoff, R, Essner, R, Nieweg, OE, Roses, DF, Hoekstra, HJ, Karakousis, CP, et al. (2006). Sentinel-node biopsy or nodal observation in melanoma. *New England Journal of Medicine* 355, 1307–1317.

Parast, L, Cheng, SC, and Cai, T (2011). Incorporating short-term outcome information to predict long-term survival with discrete markers. *Biometrical Journal* 53, 294–307.

Parast, L, Cheng, SC, and Cai, T (2012). Landmark prediction of long-term survival incorporating short-term event time information. *Journal of the American Statistical Association* 107, 1492–1501.

Park, T, Jeong, JH, and Lee, JW (2012). Bayesian nonparametric inference on quantile residual life function: Application to breast cancer data. *Statistics in Medicine* 31, 1972–1985.

Peterson Jr, AV (1977). Expressing the Kaplan-Meier estimator as a function of empirical subsurvival functions. *Journal of the American Statistical Association* 72, 854–858.

R Development Core Team (2012). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: http://www.R-project.org/.

Ries, LAG, Reichman, ME, Lewis, DR, Hankey, BF, and Edwards, BK (2003). Cancer survival and incidence from the Surveillance, Epidemiology, and End Results (SEER) program. *The Oncologist* 8, 541–552.

SAS Institute Inc. (2008). *SAS/STAT User's Guide, Version 9.2*. SAS Institute. North Carolina.

Schwarz, G (1978). Estimating the dimension of a model. *The Annals of Statistics* 6, 461–464.

SEER (2008). *Surveillance, Epidemiology, and End Results Program Research Data (1973-2008)*. www.seer.cancer.gov. National Cancer Institute, DCCPS, Surveillance Research Program, Surveillance Systems Branch, released April 2011, based on the November 2008 submission.

Serfling, R (2011). Asymptotic relative efficiency in estimation. In: *International Encyclopedia of Statistical Science*. Springer, 68–72.

Shanno, DF (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of Computation* 24, 647–656.

Shanno, DF and Kettler, PC (1970). Optimal conditioning of quasi-Newton methods. *Mathematics of Computation* 24, 657–664.

Shults, J, Sun, W, Tu, X, and Amsterdam, J (2006). On the violation of bounds for the correlation in generalized estimating equation analyses of binary data from longitudinal trials. *UPenn Biostatistics Working Papers.* Working Paper 8. `http://biostats.bepress.com/upennbiostat/art8`.

Solis-Trapala, IL and Farewell, VT (2005). Regression analysis of overdispersed correlated count data with subject specific covariates. *Statistics in Medicine* 24, 2557–2575.

StataCorp LP (2013). *STATA Multilevel Mixed-Effects Reference Manual 13*. Stata Press. College Station, TX.

Thall, PF and Vail, SC (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 657–671.

Vollmer, RT (2008). Tumor length in prostate cancer. *American Journal of Clinical Pathology* 130, 77–82.

Wald, A (1943). Tests of statistical hypotheses concerning several parameters when the number of observations is large. *Transactions of the American Mathematical Society* 54, 425–482.

Wang, BY, Goan, YG, Hsu, PK, Hsu, WH, and Wu, YC (2011a). Tumor length as a prognostic factor in esophageal squamous cell carcinoma. *The Annals of Thoracic Surgery* 91, 887–893.

Wang, SJ, Wissel, AR, Luh, JY, Fuller, CD, Kalpathy-Cramer, J, and Thomas Jr, CR (2011b). An interactive tool for individualized estimation of conditional survival in rectal cancer. *Annals of Surgical Oncology* 18, 1547–1552.

Winkelmann, R (2004). Health care reform and the number of doctor visits – an econometric analysis. *Journal of Applied Econometrics* 19, 455–472.

Xing, Y, Chang, GJ, Hu, CY, Askew, RL, Ross, MI, Gershenwald, JE, Lee, JE, Mansfield, PF, Lucci, A, and Cormier, JN (2010). Conditional survival estimates improve over time for patients with advanced melanoma. *Cancer* 116, 2234–2241.

Xu, R and O'Quigley, J (2000). Proportional hazards estimate of the conditional survival function. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 62, 667–680.